Reference, Existence and Truth in Discourse.

Catriona McPherson

PhD

University of Edinburgh

1996

Acknowledgements.

The first and biggest share of my gratitude belongs to my heroic supervisor Ronnie Cann, who has been inspirational, encouraging, challenging, critical, understanding and very, very patient. I cannot imagine having had to wipe more tears of laughter from my eyes during discussions of NP semantics with anyone else.

Thanks too to the following, who have been generous with their time, advice and expertise at various points during the production of the thesis: Jim Hurford, Louise Kelly, Chris Lyons, Irene McLeod, Cedric McMartin (sorry, Ced), Jim Miller, Etsuko Oishi, Massimo Poesio, Dan Robertson, Ludovica Serratrice and Magda Stroinska.

If all else had failed I would have got this thesis finished thanks to Ethel Jack, since the sting of Ethel's unique style of encouragement can only be borne for so long. I am grateful to her for this and for the inspiration of her constant courage and cheerfulness.

Thanks are due to my ex-roommate Simon Kirby for giving me someone to feel calmer than, and my new roommates Etsuko Oishi and Ineke Mennen for putting up with me and pretending not to notice the recent disintegration of my personality.

Something more than thanks is due to Louise Kelly and Diane Nelson for all kinds of support, for Di's cooking, and for Lou's impressive range of pep-talks.

I am grateful to the rest of the staff, students and lunchtime guests in the Linguistics Department at Edinburgh just for being around and showing me (even recently) that life can be fun.

My parents, Jim and Jean McPherson, have been wonderful during my second tour as a student. I would like to thank them for putting me up, listening to my panics and grumbles, and never believing me when I try to prepare them for my imminent failure.

I would like to thank my big sisters and their families for providing three havens where I could forget semantics and get back to eating garlic bread, drinking wine and bathing babies; thanks especially to Sheila for the time-share option on her sofa-bed, to Audrey for the PC, and to Wendy for the summer of '93.

Many thanks to my God-mother Doreen McPherson who provided the perfect solution to over-work when she introduced me to the joy of soaps.

I am especially grateful to the friendship and understanding during the last five years of Nancy and Jeff Balfour and Catherine and Olivier Lepreux who were bound neither by shared suffering nor blood-ties to care about how my thesis was going, but who have been four bricks nevertheless.

Finally, I suppose I should thank Neil McRoberts who hates effusive acknowledgements but who has been more patient than not and quite helpful sometimes.

This is for Neil with love and thanks.

Abstract.

It is a long established and still respectable claim in the linguistic discipline that sentences containing reference to non-existent objects have no truth-value. This thesis is an attempt to provide a richer and more accurate account of the interesting relations which connect the existence or non-existence of objects and the expressions by which speakers attempt to refer to them with the range of truth-values assigned to sentences containing such expressions.

After an introductory chapter which defines the main terms used in the thesis and discusses preliminary issues, the second chapter is taken up with a critical review of the history of presupposition in linguistics. The important early theoretical contributions are surveyed and the relevant later theories discussed, particularly those analyses which acknowledge the complexities of the relationship between failed presuppositions and truth-values. These are evaluated with respect to their empirical and theoretical adequacy and the chapter concludes with a summary of the outstanding problems.

Chapter three contains a discussion of determiners focusing on the existential aspects of their meaning. An analysis of existential force as a scalar phenomenon is proposed and examples of each of the types of determiner distinguished by the proposed existential scale are examined.

In the fourth chapter, questions surrounding the nature of the existence of objects are discussed. It is proposed that different types of existence are viewed in terms of different existential locations; and that these existential locations can be modelled as possible worlds. Some preliminary questions concerning the nature of possible worlds are addressed and arguments from the linguistic and philosophical literature in favour of viewing possible worlds as existent entities or as abstract constructs are reviewed. The chapter concludes by defending the modal realistic stance which maintains that all possible worlds actually exist.

Chapter five contains a detailed discussion of the members of the set of possible worlds and describes a rich structure which can be imposed on the set in the form of accessibility

relations of several types: counterpart relations which define inter-world proximity; temporal links which can be used to identify distinct temporal stages of worlds as chronological counterparts; and familiarity relations which connect individuals to sub-sets of worlds via epistemic links.

The next chapter presents a theory of how referring expressions are interpreted by means of locating their referents in possible worlds. The processes involved in accessing worlds of all types are discussed and then a set of rules is laid out which governs the choice of referential location on the basis of the relative accessibility of competing potential reference worlds, where accessibility is determined by the structure of links imposed on the set. The application of the rules to a range of increasingly complex sentences is discussed .

The final chapter of the thesis is concerned with the potential problems and wider implications associated with adopting the location theory of reference presented. First, it is shown that two related types of sentences, those containing deictic pronouns and names, which were problem cases for previous theories, can be accounted for. We then discuss the implications, concerning the failure of Leibniz's Law, which arise from treating all referring expressions as picking out their extensions. Finally there is a discussion of the implications of adopting the framework set out in the current thesis for the interpretation of a wider range of predicate types and further areas of research are suggested.

Table of Contents

Chapter 1
Introduction

1.1 General Introduction.

This thesis is concerned with the linguistic relations between reference, existence and truth. These three properties usually come together in the linguistic and philosophical literature in relation to the concept of presupposition.

The central tenet of the doctrine of semantic presupposition is that the occurrence in some sentence of a definite description whose intended referent does not exist results in truth-valuelessness for both the presupposing sentence and its negative counterpart. For example, the theory holds that sentence 1.1. is associated with the existential presupposition 1.2. and that if 1.2. is false (if there is no gravy) then both 1.1. and its negative counterpart 1.3. are neither true nor false.

1.1. The gravy congealed.

1.2. The gravy exists.

1.3. The gravy did not congeal.

There are many clear examples which seem to bear out the claims of presuppositional theories very well. Sentence 1.4., for example, contains a definite description whose referent is easy to accept as non-existent, and given a choice between *true*, *false* and *neither* to describe this sentence, linguistically naive native speakers do categorise it in the third way. In contrast, such sentences as 1.5 are perceived as true by many speakers and those such as 1.6 are perceived as false.

1.4. The King of France is bald.

1.5. The Prince of Wales is married.

1.6. The Queen of Spain is Welsh.

The problem is that not all apparent instances of presupposition failure behave in the same way. It is possible to identify many sentences which are less clear cut, as well as a number which appear to be counter-examples. Sentence 1.7. *should* be perceived as a truth-valueless statement just as 1.4. was, given that dodos no longer exist but it is equally possible to analyse sentence 1.7. as true.

1.7. The dodo is a flightless bird.

Sentence 1.8. is an even less convincing example of the effect of presuppositional failure.

1.8. The dodo cannot fly.

It is possible to argue that 1.8. is true simply because 1.7. is true and they are paraphrases of each other. More importantly, it is possible to argue that 1.8. is true because 1.9. is false.

1.9. The dodo exists.

On presuppositional theories, the falsity of 1.9. should ensure the *truth-valuelessness* of 1.8. If the falsity of 1.9. allows us to deduce that 1.8. is true (if we reason that dodos cannot fly because they do not exist) then 1.8 is simply an entailment of the negation of 1.9 and there is no need to invoke the special category of presupposition. It should be pointed out that sentences 1.7 - 1.9 contain a definite description with generic reference. However, since generic reference is a common use of such expressions, a theory such as the theory of presupposition, which purports to account for definites, ought to be able to handle generic uses.

Not all apparent counter-examples to presuppositional theories are so straightforwardly handled however. In particular, it appears that some non-existent entities are more non-existent than others. Further examples containing existentially questionable referents are easy to find. For example,

1.10. The starship Enterprise is powered by a warp drive.

1.11. The tooth-fairy used to deal in sixpences but now leaves pound coins.

Our ability to assign truth-values to such sentences indicates the need to recognise a range of existential possibilities which includes "existence in the real world" and the classical "non-existence" of the King of France but also includes such categories as the "fictional existence" in 1.10 and the "mythological existence" in 1.11.

There is a further set of problem sentences, of which 1.12 is an example.

1.12. The King of France is standing next to me.

1.12, since it contains the archetypal definite description with a failed presupposition, "should" be truth-valueless, but it is even more clear in this case that this sentence *must* be analysed as false, since a non-existent entity cannot stand in a spatial relation such as proximity to an existent one. It is not at all clear that it is the possibility of arguing for the existence of the King of France that allows this analysis. On the contrary it seems that the acceptance of the non-existence of this individual along with the acceptance of the existence of the referent of *me* is precisely what forces the assignment of the value *false*.

Clearly there *is* a connection between non-existent referents and truth-valuelessness. However, it appears that not all instances of non-existence cause truth-valuelessness, nor that all instances of truth-valuelessness are caused by non-existence. The problems associated with presupposition failure are solvable only if they are seen as part of a larger issue concerning *all* means of reference and *all* types of existence. It is necessary to work from a different perspective than that taken by presuppositional analyses, on a variety of parameters, to allow a fuller and more accurate account of this interesting area.

For the remainder of this chapter, we will lay out the means by which we intend to broaden

the focus away from the traditional concerns covered by presupposition analyses. A detailed discussion of presupposition itself appears in the following chapter.

1.2. Reference

Presuppositional theories as a means of accounting for the relations between reference, existence and truth suffer from a narrowness of focus in that only definite descriptions (usually singular), only one type of context, and not even all uses of definite descriptions are addressed.

There does not appear to be any a priori reason to restrict the concept of presupposition to only one NP type. We consider existential aspects of the meaning of singular and plural definites, *the King of France*, *the MPs on both sides of the House*; singular and plural indefinites, *a bugging device*, *some frozen turnips*; pronouns; names, *Zsa-Zsa Gabor*, *Mr Chekhov*; bare plurals, *ribbons*; and a range of quantified NPs, *both Poles*, *neither Equinox*, *every hamster*, *any excuse*, *few authors*, *no solution*. Not all of these expressions refer in what is the standard understanding of this term proposed by Lyons (Lyons, 1977).

> "It is a condition of successful reference that the speaker should select a referring expression - typically a proper name, a definite noun-phrase or a pronoun - which ... will enable the hearer ... to pick out the actual referent from the class of potential referents."
>
> (Lyons, 1977: 180)

We will follow Lyons in assuming that reference is a feature of expressions when they are used in particular discourse contexts, rather than a terminological variant for denotation, but we will use "reference" and "location of a referent" in relation to the interpretation of all types of NPs, not just definites, pronouns and names, since it appears that non-existence affects the interpretation of all of the expressions above. What we are aiming at in this thesis is to provide an overarching theory that goes beyond presupposition failure only associated with definite descriptions to the existential implications for all NPs from which the particular instance of definite descriptions follows.

A further innovation in this thesis is that we will not only consider the referentially transparent contexts such as that in 1.13 which presuppositional theories also address (since it is possible to argue that the definite in 1.13 can be associated with the existential presupposition in 1.14) but will discuss the occurrence of nominal expressions in a range of "referentially opaque" contexts (Russell, 1905), such as the object NPs in 1.15 - 1.19.

1.13. The MPs on both side of the house shook with terror.

1.14. The MPs on both sides of the house exist.

1.15. I thought I saw a bugging device.

1.16. I dreamed about some frozen turnips last night.

1.17. I do not believe in Zsa-Zsa Gabor.

1.18. I want to marry Mr Chekhov.

1.19 I am looking for ribbons in the toaster.

Definites in these contexts are assumed to be outside the scope of presuppositional accounts since it is argued that the failure of existential presuppositions of expressions in such contexts does not regularly result in truth-valuelessness (McCawley, 1981; Chierchia and McConnell-Ginet, 1990; Partee, ter Meulen and Wall, 1990).

The last important way in which the current approach to reference and existential aspects of meaning broadens the focus from that of traditional presuppositional theories is by addressing a wide range of the *uses* of nominal expressions and attempting to relate these uses to each other in a unified manner. We consider uses of definites which are anaphoric (Hawkins, 1978; Sperber and Wilson, 1986; Werth, 1993) and uses of definites as "first-mentions" (Hawkins 1978): for example

1.20. I had a cat and a budgie but **the cat** died.

1.21. I had a budgie but **the cat** ate it.

We compare these (and other) uses of definites to anaphoric and deictic uses of pronouns: for example

1.22. I had a cat but **she** died.

1.23. Here **she** comes now.

We also examine the uses of names, to evaluate their anaphoric potential and their appropriateness as first-mentions, compared with definites and pronouns. For example:

1.24. **?**I am fond of **my husband**$_i$, but **Brian**$_i$ has terrible clothes.

There are also interesting alternations to be considered between definite descriptions used as rigid designators (i.e. as names) and identical definite descriptions used as non-rigid designators (i.e. as definite descriptions) (Kripke, 1980). For example, it is possible to paraphrase 1.25. as 1.26 since *the Prince of Wales* and *Prince Charles* are synonymous if the definite is understood as a rigid designator:

1.25. The Prince of Wales is still married.

1.26. Prince Charles is still married.

However, it is not possible to paraphrase 1.27. as 1.28 since *the Prince of Wales* in 1.27 is not being used as a rigid designator.

1.27. The Prince of Wales is always the heir to the throne.

1.28. **?**Prince Charles is always the heir to the throne.

In addition, generic uses of singular definites and indefinites and bare plurals are considered. For example:

1.29. The signing chimp is a miracle of modern times.

1.30. A signing chimp is wondrous to behold.

1.31. Signing chimps make perfect pets.

We attempt to relate the existential aspects of the meanings of generics to other types of usage and to account for the gap in the generic pattern caused by the impossibility of a generic interpretation of a plural definite, such as that in sentence 1.32.

1.32. The signing chimps attract large grants.

A further innovation proposed in this thesis is that NPs do not necessarily entail existence or non-existence discretely. The set of quantifiers are traditionally divided into two discrete groups on the basis of a binary split between those with existential force (e.g. *both*, *neither*, *all, many*, *some*) and those with no existential force (e.g., *every*, *any*, *few*, and *no*). We will consider the availability of existential interpretations of some of the supposedly non-existential quantifiers. For example, the subject NP of the deepest subordinate clause of 1.33. can be interpreted existentially, which allows the interpretation of sentence 1.34 as true, but can also be interpreted non-existentially, which allows the interpretation of sentence 1.35. as true.

1.33. I decided to buy chocolates for everyone who helped.

1.34. I didn't grudge the expense since they had all been very kind.

1.35. This cost me nothing as nobody had bothered to help at all.

We will account for this range of expressions and uses in a unified way by placing them at ordered points on a scale of existential strength. By proposing such a scale we do away with the notion that existential force is a property with a pair of binary values. This departure is mirrored in our treatment of the existential properties of entities which, we will argue, consist of a richer range of possibilities than just existence or non-existence.

1.3. Existence.

The examples given as Nos. 1.04 - 1.11 in section 1.1. were intended to indicate that a very narrow binary definition of the existence of entities is not adequate to account for the existential possibilities associated with the referents of nominal expressions. If existence is defined as existence in the real world, or even in some contextually salient sub-part of the real world, then only the first and last of the definite descriptions in 1.36 are clear and reliable cases of either existence or non-existence.

1.36. The highest mountain in Scotland.
The last day of the miners' strike.
The children of William the Fifth.
The Pope's wife.
The original Baker Street Irregulars.
The oldest surviving Leprechaun.
The squared circle.

The intervening expressions can be seen as examples of other types of "existingness" which are neither existence nor non-existence. The first illustrates past existence in the real world, and the others, probable future existence in the real world, possible future existence in the real world, existence in a fictional world, and existence in a mythological world.

It is certainly not the case that all sentences containing any of the "existentially questionable" expressions in 1.36 are perceived as truth-valueless. For example, it is necessary to interpret sentence 1.37 as a true statement, since it is analytic.

1.37. The oldest surviving Leprechaun is a Leprechaun.

However, in what does the analyticity lie? The subject NP must be taken as referential and not generic because of the use of the superlative. A generic interpretation involving universal

quantification will necessarily be true, whereas its existential or referential counterpart would be truth-valueless on presuppositional theories.

One way to account for this is to say that although Leprechauns do not exist in the real world, in those mythological worlds in which Leprechauns do exist, they are Leprechauns. In the same way, we can account for the intuitive truth of such sentences as 1.38.

1.38. The King of France is a King.

We can say that sentence 1.38 is true because in every world in which the King of France exists he is a king. Our knowledge that Leprechauns are Leprechauns and kings are kings, means that if we can accept the existence of these entities in other possible worlds we can assign the value true to the sentences. Similarly if we can accept that the referents of the subject NPs in sentences 1.39 and 1.40 exist in other possible worlds, we can use our knowledge about these worlds to analyse sentence 1.39 as true and 1.40 as false.

1.39. The starship Enterprise is powered by a warp drive.

1.40. The tooth-fairy comes down the chimney on Christmas Eve.

However, by taking this route to an explanation of the truth-values of 1.37 - 1.40, we render ourselves less able to explain the intuitive truth-valuelessness of sentences such as 1.41.

1.41. The King of France had a perm.

If we allow that the King of France exists somewhere, in some possible world, it is not coherent to appeal to the notion of his *non*-existence to account for the truth-valuelessness of the sentence. The problem with employing the notion of possible worlds is that if we recognise the existence of any one world besides the real world we are committed to accepting the existence of every possible world. This entails that for any sentence there is at least one world (and probably an infinite number of worlds) in which that sentence is true and at least one world in which that sentence is false. There is no valid argument for allowing worlds

where the King of France exists but disallowing worlds in which the King of France exists and has a perm.

The problem we must address is whether we can accept the existence of every possible world, by which means we are provided with a potential existential location for the referent of any nominal expression, but at the same time constrain the *use* of possible worlds in interpretation, so that all referring expressions are matched with referents at locations around the set in such a way as to reflect our intuitions about the truth-values of the sentences in which they occur. This task depends on a particular view of truth and truth value assignment, and establishing the validity of this view is the final piece of preliminary argument which must be carried out before embarking on the development of the argument in detail in the main body of the thesis.

1.4. Truth.

If we disregard the need to account for presuppositional effects, it is possible to argue, in accordance with classical bivalent logic, that every well-formed sentence is associated with one of two truth-values: either true or false [this being a central tenet of both Montague Grammar (Dowty, Wall & Peters, 1981) and DRT (Kamp & Reyle, 1993)] However, only DRT sees this as a failing in the theory, as illustrated by the following quotation:

> "It may well be... [of some sentences] ...that all that can be ascertained of them is that they are true in some situations and false in certain others, but that there are many other situations in which their truth values are not determined."
>
> (Kamp & Reyle, 1993, p:469)

However, if we seek to account for presuppositions in semantic terms, it seems that we cannot maintain this ideal state of affairs and must choose one of two alternative positions: either that there are three truth-values or that there are two truth-values and the possibility exists that sentences are associated with truth-value gaps. Burton-Roberts (1989) discusses these alternatives and reviews the opposing arguments to be found in the literature, including Keenan's (1972; 1973) development of three valued logic, Martin's (1975) two-dimensional

four-valued logic, and the various suggestions by van Fraassen (1969), Rescher (1969), Herzberger (1970) and Seuren (1985), among others, that a system of two truth values with truth value gaps is to be preferred; a view which Burton-Roberts (1989) supports. He notes for example, Rescher's view that:

> " the further a system departs from orthodox logic...the more tenuous its claim to constitute a "logic" will be."
>
> (Rescher, 1969: p129)

As well as this general reservation about deviations from classical logic, there is a second motivation of a very different kind, for adopting the "2-values with gaps" approach. This arises from our desire to reflect intuitions and model as accurately as possible the procedures which occur during discourse. Briefly, if there are three truth-values of equal status it is problematical to reflect the notion that one of these, the third truth-value, is indicative of a problem, possibly a breakdown, in the normal functioning of a discourse. If on the other hand, we have truth-value gaps where presuppositions fail, it is a simpler matter to label these gaps, like other unfinished interpretation tasks, as evidence of failures in the discourse.

Furthermore, it is possible keep the notion of a simple bivalent logic under which *every* well-formed sentence must have a truth-value, and at the same time allow the non-assignment of truth-values to utterances of sentences in discourse contexts.

To do this we must accept a particular view of what Strawson called "aboutness" (Strawson, 1964). If we accept that statements are about entities and not about worlds as is sometimes claimed (McCawley, 1981) then a sentence such as 1.42. must be about an individual who is the King of France and only indirectly about a place (a world) where he exists.

1.42. The King of France is very light on his feet.

If this sentence is perceived as truth-valueless (as we assume it is) then it might seem that the world where this individual, the King of France, exists has very different properties from the

real world. In the real world it is either true or false of every individual (or at least of every human individual) that he or she is light on his or her feet[1]. This might mean that to accept that sentence 1.42 is truth-valueless we must accept that the world in which the King of France exists is incomplete[2].

However, if we can accept that truth-value assignment is a firmly pragmatic phenomenon there is another possible explanation. By accepting that truth-value assignment is a process carried out by real hearers with respect to utterances in real discourse contexts, it becomes unremarkable to claim that it is often not possible for truth-value assignment to proceed in an ideal fashion.. For example, given the current discourse context, where the current author is in the role of speaker and the current reader is in the role of hearer, 1.43. is the representation of a sentence for which no truth-value assignment is possible

1.43. The daughter of my eldest sister has no middle name.

We do not have to analyse this sentence as a statement about an entity in a non-real world with the mysterious property of internal incompleteness to account for the inability of the current reader to assign a truth-value to it. We can just as easily argue that nothing prevents the assignment of a truth-value except the limits on the epistemic access of the current reader to the relevant details of the real world.

We will argue in chapter 4 that we can quite readily extend this argument to account for sentences such as 1.42: it may not be possible for a hearer of 1.42. to assign a truth-value if he or she does not have the necessary epistemic access to the world where the referent exists. This argument is entirely unconnected to any claim that in any world where the King of France exists the individual is in some third state which is impossible in the real world. Likewise, the argument is compatible with the claim that truth or falsity are unavoidable

---

[1]The possibility for fuzzy interpretations of "lightness on one's feet" does not contradict this claim. If an individual is neither light on their feet nor heavy on their feet all this means is that they have the property of lightness on their feet to 50% and the opposite property to 50%. The "all-light" and the "all-heavy" options are simply two extreme cases. (Russell, 1923; Zimmerman, 1985).
[2]For an extended review of the arguments for and against incompleteness see Chapter 4 section 4.4.2. and Chapter 5 section 5.3.1.

entailments of well-formed sentences.

In summary, we will claim later in the thesis that truth-valuelessness is a possible relation between an utterance, a proposition and a hearer, but that truth-valuelessness is not a possible relation between a proposition and a sentence. There is no equivalent distinction which can be applied to tokens of written sentences such as the distinction between *utterance* and *sentence* in spoken contexts. In the present written format therefore, we are forced to talk about the truth-values of sentences but this should always be understood as a reference to the truth-value of a written sentence token. To place the introductory arguments, which we have outlined above, in the context of the thesis as a whole, this chapter concludes with a brief outline of the development of the argument in the remaining chapters of the thesis.

1.5. Chapter Outline.

Chapter Two contains a review of the history of presupposition in the field of linguistics. The important early theoretical contributions (Frege, 1981; Russell, 1905; Strawson, 1950) are discussed and there is a brief account of the large literature which followed (e.g. Morgan, 1969; Fillmore, 1971; Kiparsky and Kiparsky, 1971; Gazdar, 1979; Karttunen, 1974). Particular attention is paid to three approaches to the area, by Fodor (1979), Burton-Roberts (1989) and Werth (1993), which provide some of the theoretical basis on which the analysis to be presented in this thesis is built. There is a discussion of three perennial problem areas associated with presuppositional accounts, namely cancellability, ambiguous negation and the influence of context. The chapter concludes with an evaluation of the argument put forward in Werth (1993) for the abandonment of the category of presupposition altogether.

The following chapter is concerned with the existential features of the meaning of nominal expressions. We attempt to provide a review of the full range of nominal expressions: pronouns, names, definites, indefinites and quantified NPs and to address all the uses of these: anaphoric, assertive, deictic and generic. We also discuss the occurrence of referring expressions in opaque contexts. Finally, we propose a scale of existential strength on which

all types of nominals can be placed.

Chapter Four turns from the existential features of linguistic expressions to consider the existential possibilities for entities. We propose viewing different types of existence as different existential locations and introduce the notion of modelling existential locations using a set of possible worlds. We review competing definitions of possible worlds either as abstract constructs (Bradley and Swartz, 1979; McCawley, 1981) or existent entities (Bradley and Swartz, 1979; Lewis, 1986). An argument for a modal realistic view of the nature of possible worlds is presented. In the later sections of this chapter we introduce the notion of the correct choice of reference world as a factor in the proper interpretation of nominal expressions and offer a short initial argument which presents evidence in favour of the usefulness of the emerging location theory.

In the fifth chapter we define the members of the set of possible worlds and describe the arrangement of the set. We distinguish the current, real world and relate this to its temporal counterparts using a type of accessibility relation which is constrained by a principle of consistency. We propose a more general type of accessibility relation which connects other counterparts and discuss the question of whether there are any limits to the set of possible counterparts to the real world; i.e. whether there are such things as impossible worlds. Connected to this question is the matter of the possibility of incomplete or incompletely specified worlds. Having defined the arrangement of the set of possible worlds in terms of the relative closeness of counterparts, modelled in terms of accessibility relations, we propose a final layer of structure which is expressed as a set of familiarity relations which hold between individuals in some world and some sub-set of other worlds with which they have epistemic links.

Chapter Six is concerned with the process of interpretation. First we discuss some preliminary issues surrounding our view of interpretation; the distinction between the interpretation of referring expressions and the interpretation of utterances as a whole, the validity of viewing interpretation as truth-value assignment and the question of the type of linguistic construct to which truth is assigned. We then propose a set of rules for

interpretation using the structured set of possible worlds. This amounts to a group of rules and constraints which governs movement around the set and predicts which world is chosen as the existential location for each referring expression. By working through increasingly complex examples in detail we show that the rules predict the correct truth-values for a range of sentences.

The final chapter addresses the wider implications of the theory proposed; both further applications and potential problems. We return to particular problems with the interpretation of pronouns and names identified by Fodor (1979). The current theory allows the difficulties Fodor identified to be stated as one coherent problem; which can be summarised as the fact that pronouns and names do not seem to conform to what we argue is the general rule with respect to triggering movement around the set of possible worlds. We propose an explanation which relies on the acceptance of the scalar ordering of existential force. We also discuss the implication of the proposed view of interpretation for sentences with nominal expressions in opaque contexts and attempt to reconcile the location theory with the failure of Leibniz's Law. Finally, there is a discussion of the need for further technical refinements in the expression of the structure of the set of possible worlds to allow the model of temporal counterparts to handle a greater variety of predicate types, especially accomplishments and habitual actions.

Chapter 2.
Presupposition

2.1. History of Presupposition in Linguistics.

There is a vast amount of linguistic literature surrounding the areas of presupposition and presupposition failure in which we can identify three important and enduring landmarks. All debate about presupposition in the field of linguistics stems from the philosopher Frege's discussion of referring expressions in "On Sense and Reference" (Geach and Black, 1980). However after Frege, Russell's Theory of Descriptions (Russell, 1905) which did not permit the notion of presupposition held sway for almost fifty years until Strawson's famous attack (Strawson, 1954) which led to the modern upsurge of interest. These three works are discussed below. However, since much of the subsequent presuppositional analyses[1] are irrelevant from the point of view of the current thesis they will, to a large extent, be ignored although we will review the problem areas which led to much of the disagreement; i.e. the problem of cancellability and the ambiguity of negation. We will concentrate on three modern theories concerning related areas: J.D. Fodor's analysis of cross-world reference (Fodor, 1979) which provides the foundation for a good part of the theory presented here, Burton-Roberts proposal for the retention of the category of presupposition as part of semantics rather than pragmatics (Burton-Roberts, 1989) and Werth's argument for the abandonment of the category of presupposition altogether (Werth, 1993).

2.1.1. Frege's Theory of Presupposition.

Frege stated that any proper name used in an assertion has reference. He argued that this feature of proper names should be viewed as presupposition, which is constant under negation, rather than as part of the meaning of the assertion.

---

[1] in particular, most of the work produced in the 1970's which was concerned with sentence level explanations of the cancellability of presuppositions in various contexts, (e.g. Kiparsky and Kiparsky, 1971; Karttunen, 1973; 1974; Wilson, 1975; Katz & Langendoen, 1976; Gazdar, 1979). This work has been reviewed from different perspectives by Kempson (1977) and Levinson (1983) and its relevance to the current thesis is discussed in section 2.2.2.

> "If one . . . asserts 'Kepler died in misery' there is a presupposition that the name Kepler designates something. . . . That the name Kepler designates something is just as much a presupposition of the assertion 'Kepler died in misery' as for the contrary assertion.
>
> (Frege, 1981: 69-70)

Frege's opinions on the result of presupposition failure are to be found in his discussion of temporal phrases which he treats similarly to names. He observes that the "sense" of the sentence reproduced here as 2.01 could be rendered in the paraphrase given here as 2.02.

2.01. After Schleswig-Holstein was separated from Denmark, Prussia and Austria quarrelled.

2.02. After the separation of Schleswig-Holstein from Denmark, Prussia and Austria quarrelled.

He states that in 2.02 the proposition that Schleswig-Holstein was separated from Denmark is not being asserted but is a necessary presupposition for the subordinate clause to have any reference[2]. In contrast, he proposes that in 2.01 the subordinate clause may constitute either an assertion or a presupposition. He argues that if some hearer believes it to be false that Schleswig-Holstein was once separated from Denmark he or she will take 2.01 on the first reading, where the subordinate clause is an assertion, to contain a thought which is false (the subordinate clause) and a part with no reference (the main clause). On the other reading, where the subordinate clause is a presupposition, the same hearer, according to Frege, would take 2.01 as a whole to be without a truth-value because of a lack of reference for its subordinate clause. Likewise, he argues that a hearer ignorant of the historical facts will take 2.01 (on either reading) or 2.02 to be wholly without reference.

The major frustration for any objective discussion of Frege's work is that his style of writing and use of terminology leave his proposals open to many different interpretations. It is all too easy to find in his work whatever you want most to find there. And in relation to

---

[2]on the assumption that stretches of time are logical objects.

presupposition this is very much the case since his discussion is short and somewhat vague. It is tempting, given our interest in the area, to speculate that Frege distinguished between the lack of a truth-value and lack of reference but we cannot be sure that this inference is justified. However, it does seem plausible to infer from Frege's argument that he believed that the failure of an assertion or presupposition can be seen as a function of the state of knowledge of a hearer as well as a function of the inherent truth or falsity of the presupposed or asserted proposition given that he takes account of the belief state of the hearer in his discussion of the interpretation of 2.01 and 2.02. Frege states that a hearer who is ignorant of the facts of European history will take sentences such as 2.01 and 2.02 to be "without reference". We can infer from this that he saw no difficulty in the idea that a statement might have a truth-value discernible by one hearer but that the same statement in the same context might be perceived as truth-valueless by some other hearer. This view of truth-value assignment, which defines truth-valuelessness with respect to individual hearers, has either been ignored (Strawson, 1950) or rejected (Burton-Roberts, 1989) by later proponents of presupposition. We will discuss this approach to questions of truth-valuelessness in detail in section 2.2 of the current chapter.

2.1.2. Russell's Theory of Descriptions.

Russell's theory of descriptions (Russell 1905), although it constitutes a direct contradiction of Frege's approach, was actually born out of Russell's dissatisfaction with other earlier theories of the meaning of (in)definite descriptions; e.g. that of Meinong (1904), which treated them as simple referring expressions directly denoting individuals in the world. The Meinongian concept of a referring expression captured naive intuitions about how some such expressions might work but could not deal in an adequate way with any but the most simple cases such as **the sun**, or **my father** which a Russellian account might seem to complicate too much. However, a direct denotation account runs into difficulties in the analysis of such expressions as **fewer cats,** or **either option** since it is unclear what objects such expressions might denote. A theory which takes definites and indefinites to be direct referring expressions would have to provide a separate account for quantified NPs, as in DRT (Kamp and Reyle, 1993).

A much more central objection of Russell's was that such a theory could not account for the availability of meanings for NPs which refer to non-existent objects. Russell argued that if the meaning of an expression is the object it refers to then an expression such as **the King of France** should be meaningless if France has no King and that since **the King of France** is comprehensible although France has no King, then the meaning of a definite description cannot be its reference.

Russell proposed instead that only names or rigid designators actually refer and that definite and indefinite descriptions should be analysed instead as expressions of existential quantification which are similar to other quantified expressions and whose semantics, like the semantics of other quantified expressions, could be explained in set-theoretical terms. An indefinite description such as **a cat** has the logical form $\exists x$ *(cat (x))* and its meaning can be expressed as a statement that *the set of all things which are cats has at least one member*. A sentence containing this indefinite description such as 2.03 has the logical form given in 2.03a. and the meaning given in 2.03b.

2.03 **A cat yowled.**

2.03a. $\exists$x (cat(x) & yowled(x))

2.03b. *the intersection of the set of all things which are cats and the set of all things which yowled has at least one member*

The translation and interpretation of a definite, for Russell, is identical to that of an indefinite as regards existential quantification, but in addition, Russell argued that a definite carries a uniqueness entailment. For example **the Pope** has the logical form $\exists x$ *(pope (x) &* ($\forall y$ *(pope(y)* $\rightarrow$ *y = x)))* and the resulting interpretation can be expressed by the statement *the set of all things which are pope has exactly one member*. A sentence such as 2.04 containing a definite description has the logical form given in 2.04a and the meaning given in 2.04b.

2.04. **The Pope wears a funny hat.**

2.04a. $\exists x$ (pope (x) & ($\forall y$ (pope(y) $\rightarrow$ y = x)) & wears a funny hat (x)).

2.04b. *The intersection of the set of all things which are popes and wear funny hats has exactly one member*

On a Russellian account, 2.04 is true if and only if there is one pope who wears a funny hat and is false if either there is no pope, or there is more than one pope, or the one and only pope does not wear a funny hat.

Russell's account of indefinites provides a satisfactory account of negated indefinites. Since in his theory an indefinite such as **a black cat** does not refer to any individual, a sentence such as 2.05, which has the logical form given in 2.05a constitutes an assertion, given in 2.05b, that nothing that is both a cat and black crossed some path.

2.05. **A black cat didn't cross my path.**

2.05a. ~ ($\exists x$ (cat (x) & black (x) & crossed my path (x))

2.05b. *the intersection of the sets of cats, black things and things that crossed some path is empty*

According to this theory, 2.05 is equivalent in logical structure and semantic content with 2.06 and such an equivalence seems to accord perfectly with our intuitions as native speakers about these two sentences.

206. **No black cat crossed my path**

The analysis thus accounts for the perceived anomaly of such a sentence as 2.07.

2.07. **A black cat didn't cross my path, but another one did**.

Russell's proposals about the interpretation of negated definites, however, are less satisfactory. Sentence 2.08 has the (simplified) logical form given in 2.08a and the meaning

given in 2.08b.

2.08. **The Pope's wife does not wear a funny hat.**

2.08a. ~ (∃x (pope's wife (x) & ∀y (pope's wife (y) → y = x) & (wears a funny hat (x)))

2.08b. *It is not the case that the intersection of the set of all things which are the pope's wife and the set of all things which wear funny hats has exactly one member.*

The result of a Russellian analysis of the meaning of 2.08 is that it is predicted to express a true proposition about the real world because in the real world there is no such individual as the pope's wife. This does not accord with the intuitions of many speakers about such sentences, which tend to be perceived as truth-valueless or in some other way unacceptable.

2.1.3. Strawson's Attack on Russell's Theory of Descriptions.

The unintuitive interpretations of definites which were the consequence of Russell's theory of descriptions were attacked by Strawson in his paper *On Referring* (Strawson, 1950). Russell had argued that the requirements for existence and uniqueness on some singular definite expression should be seen as assertions of exactly the same status as the predicate with which that expression occurred. Strawson argued that only the proposition expressed by the predicate should be seen as an assertion. For example, while Russell claimed that 2.09a, 2.09b and 2.09c are all asserted when 2.09 is uttered, Strawson argued that only 2.09c is asserted.

2.09. The butler finished the gin.

2.09a. ∃x (butler (x)

2.09b. ∀y (butler (y) → (y = x))

2.09c. finished the gin (x)).

Strawson, like Frege, suggested that the proposition that there is a unique butler should be recognised as a presupposition of 2.09, although, unlike Frege, Strawson made it clear that he considered presuppositions to be properties of sentence tokens. Strawson proposed, again unlike Frege, that the failure of a presupposition, brought about by the falsity of the proposition it contained, resulted in the presupposing sentence being without a truth-value. Strawson argued that presuppositions were unaffected by negation, so that a sentence and its negative counterpart always shared presuppositions. A sentence such as 2.10, on Strawson's account, had the logical form given in 2.10a where only the asserted proposition is negated.

2.10. The butler didn't finish the gin.

2.10a. $\exists x$ (butler (x) & $\forall y$ (butler (y) $\rightarrow$ (y = x)) & ~ (finished the gin (x))).

Strawson's formulation of the theory of presuppositions does better than Russell's theory of descriptions in reflecting intuitions about the effect of the non-existence of referents on truth-values and the necessary conditions for the truth of a negative statement, but it raises other problems of its own. Two of these problems in particular have been the subject of an overwhelming amount of interest from linguists in the last forty years. These are the problem of presupposition cancellability and the problem of the ambiguity of negation both of which can be seen as part of a larger problem concerning context.

2.2. Problems with Presuppositional Theories

Strawson's original formulation of the theory of presupposition held that if a presupposition fails the presupposing sentence is *always* neither true nor false. However, there is a range of examples of negative sentences which are perceived as true even though they contain presuppositions which do not express true propositions. The clearest examples are those sentences in which it is explicitly asserted that the propositions associated with the presuppositions do not hold. Sentence 2.11 is an example of this type of sentence which, it has been argued[3], shows that Strawson's formulation cannot be correct.

---

[3]by Kempson (1977) for example

2.11. It's not true that my children are monsters because I don't have any children.

Sentence 2.11 can only be interpreted as true if **my children** has no reference, which is in direct contradiction with Strawson's argument that the occurrence in any sentence of the expression **my children** without any reference necessarily causes the sentence to be without a truth-value. Given the assumption that there is an existential presupposition in 2.11, there are two possible solutions to this problem. Either presuppositions can be assigned the property of cancellability or negation can be argued to be ambiguous. However, neither of these possible solutions is entirely satisfactory or without undesirable consequences.

2.2.1. The Problem of Cancellability.

Presupposition, as Strawson viewed it, could be seen as a special kind of entailment which was associated with a truth-function distinct from that of regular logical entailments but equally fixed. So long as this definition of presuppositions held, there was no question but that presupposition was a semantic phenomenon. But to account for the possibility of assigning a truth-value to sentences with apparent failed presuppositions it became increasingly popular to attribute to presuppositions the property of cancellability or defeasability (Levinson, 1983); this allowed potential presuppositions to be cancelled either explicitly by part of the preceding or following discourse as in 2.12 or 2.13 where presuppositions 2.12a and 2.13a are cancelled, or implicitly as a result of the wider background context as in 2.14 where presupposition 2.14a is cancelled.

2.12. Since I don't have any children, I can say that even though I have a career **my children** are not a menace to society.

2.12a. I have children.

2.13. **The family gathering** was not so bad this year as most, since we all spent Christmas in different places.

2.13a. There was a family gathering this year.

2.14. Sue died before she finished her thesis[4].

2.14a. Sue finished her thesis.

However, a truth-conditional semantic system, even one with three values or with systematic gaps caused by failed presuppositions, is still a logical system and it is not possible for any relation in a logical system to have a non-logical property such as cancellability. To assign the property of cancellability to presuppositions it was necessary to move them from the bounds of logical, truth-conditional semantics and into the realm of pragmatics[5]. This solution to the problem was widely adopted by linguists and presuppositions came to be seen as a kind of conventional implicature rather than as a kind of entailment. Pragmatic presuppositions, being outside semantics, could not be defined in terms of their effect on truth-values, and presupposition failure came to be recast in terms of appropriateness (Austin, 1962; Van Dijk, 1976; Lyons, 1977), felicity conditions (Austin, 1962; Furberg, 1971) or informativeness (Harnish, 1976; Levinson, 1983). We can say that sentences such as 2.15a. and 2.15b. are normally only appropriate, felicitous or informative when the proposition in 2.16 holds.

2.15a. The man in the moon melted.

2.15b. The man in the moon didn't melt.

2.16. There is a man in the moon.

We can also say that since appropriateness, informativeness, or felicity (henceforth we will use the term appropriateness) are pragmatic concepts without rigorously fixed definitions it is sometimes the case that sentence 2.15b is appropriately used when 2.16 does not hold. However, such explanations are unsatisfactory. If appropriateness is treated as separate from questions of truth or falsity, there is no reason for the cancellability of appropriateness conditions to be invoked to account only for such sentences as 2.15b which have failed presuppositions and not for such sentences as 2.15a. If 2.16 does not hold, sentence 2.15a cannot be *true*, but it could be used appropriately to trigger some metaphorical inference. It

---

[4]This example is taken from Levinson (1983).

[5]e.g. Karttunen & Peters (1975), Gazdar (1979).

may seem an unreasonable demand to place on a pragmatic theory to be able to account for the appropriateness of both 2.15a and 2.15b (when 2.16 does not hold) *in the same way.* However, the unreasonableness of such a demand can only be explained if we accept that presupposition failure *does* affect the core truth-values of sentences, not merely their appropriateness conditions. We can then say that, if 2.16 fails, 2.15a cannot be true although it may be appropriate if it triggers some metaphorical inference. On the other hand, if 2.16 fails, 2.15b might be true, and any theory of presupposition must be able to account for this possibility.

2.2.2. The Problem of Ambiguous Negation.

The problems associated with handling cancellability by characterising presupposition as a pragmatic phenomenon can be avoided if presupposition is treated as a semantic relation. It was proposed by some linguists in the 1970's that semantic presupposition was possible and that the problem of the truth-value of a negated sentence containing a failed presupposition could be handled by positing an ambiguous semantics for negation (Wilson, 1975, Gazdar 1979). On such an account, the first sentences in each of 2.17 and 2.18 are taken to contain two different types of negation (if the second sentences are true).

2.17. The Duke of York is not ugly. However, his clothes are terrible.

2.18. The King of France is not bald. France is a republic and her President is thinning a bit, at most.

The negation in 2.17 has been described as narrow scope or internal negation (Russell, 1905), descriptive negation (Kempson, 1977), presupposition preserving negation (Kiparsky and Kiparsky, 1971). What is negated in 2.17 is not the proposition that the Duke of York exists, but the proposition that he is ugly. In contrast, the negation in 2.18, which has been described as wide scope or external negation (Russell, 1905), denial negation (Kempson,

1977), or presupposition cancelling negation (Kiparsky & Kiparsky, 1971) does not negate only the proposition that the King of France is bald, but the entire conjoined proposition that there is a unique King of France and that that individual is bald.

It is now widely accepted that an analysis of negation as ambiguous is not well-motivated. Kempson (1977) provides a strong argument against the analysis. She points out that, if existential and predicate propositions in sentences like 2.17 are represented, as they were in Russell (1905), as a conjunction, then negation is semantically vague rather than ambiguous, since the falsity of any of the conjoined propositions is simply one of the cases of the falsity of the conjunction as a whole. Kempson goes on to argue that even proposing vagueness is ad hoc since there is no cross-linguistic evidence which would support the argument that **not,** in English, happens to be vague between two or three logically distinct readings. Kempson (1977) also identifies a problem with defining the distinction between denial and descriptive negation. Negation type must be defined independently; i.e. the definition cannot depend on the distinction between preserved and cancelled presuppositions, if presupposition failure is to be handled in a non-circular way. The narrowest definition of denial negation is as a contradiction of an immediately previous explicit statement. On this definition, the negation in sentence 2.19b can be defined as denial negation.

2.19a. Your children are monsters

2.19b. My children are not monsters.

If denial definition is defined in this way, then there are many apparent counter-examples which cannot be defined as denial negation but which seem to contain cancelled existential presuppositions. Kempson (1977) uses as an example a quotation from Strawson (1950):

> "Neither Aristotelian nor Russellian rules gives the exact logic of any expression of ordinary language: for ordinary language has no exact logic".
>
> (Kempson, 1977: p.152)

She states that the sentence in this quotation cannot be defined as an a example of denial negation but is a "descriptive negative statement finalising the previous argument" which

nevertheless contains a cancelled presupposition. Kempson argues that if the concept of denial negation is extended to embrace such counter-examples as this the distinction between descriptive negation and denial negation becomes vague to the point of uselessness and cannot be used to support a semantic presuppositional analysis.

Kempson's objections, like those of others (Lyons, 1977; Levinson, 1983) who attack the notion that negation is ambiguous, are compelling and the problems associated with positing ambiguity in negation have led to a large-scale abandonment of the search for a semantic theory of presupposition[6] in preference for pragmatic theories and the concomitant difficulties of cancellability (Lakoff, 1972; Stalnaker, 1974; Kempson, 1975; Wilson, 1975). We will now review the problems which arise from attempting to incorporate considerations of context into sentence level theories.

2.2.3. The Problem of Context.

Most past and current theories of presupposition, like many theories which address other semantic and pragmatic phenomena, treat isolated sentences as the basic unit of study. Nevertheless, since semantic and pragmatic presuppositional accounts typically make reference to such concepts as the distinction between given and new information, topic/comment distinctions, and previous statements, to account for presupposition cancelling, "the context" serves an important explanatory function. There is a widespread tendency to fill up the "pragmatic waste-basket" with awkward bits of left-over meanings. We could say that sentence level semantic theories which appeal to context are raiding the "pragmatic toolbox" for useful bits of theory.[7] Such theories are often inconsistent in that discourse elements such as the context and the background knowledge of the discourse participants are ruled out of the framework of the theory and yet are appealed to ad hoc to explain potential anomalies. For example take a recent paper by Mercer (1992):

---

[6]one notable exception being Burton-Roberts (1989) which is discussed in section 2.4

[7] Werth (1993) has attacked theories which persist in attempting to give sentence-level accounts of presupposition by making use of discourse level concepts and his alternative analysis of the area covered by existential presuppositions is discussed in section 2.5.

> "We are not trying to model communication or discourse so the only statement assumed to exist in the knowledge base of the speaker is the one just communicated."
>
> p228

> "We consider presuppositions and entailments of a sentence as inferences obtained from the logical (or semantic) representation of the sentence together with other forms of background knowledge."
>
> p230

> "Given that the sentence is uttered in a conversational context which assumes some background information (for example, the meanings of words and knowledge about the world)..."
>
> p231

Another striking feature of sentence level accounts which poach discourse level apparatus is that such theories tend to characterise the relation between individual sentences and contexts back to front. Perhaps inevitably, in theories which are fundamentally concerned with isolated sentences, the presupposing sentence is taken to be the trigger for the construction of the background context.

The following two quotations from Van der Sandt's (1988) paper (*Context and Presupposition*) and Heim's (1982) thesis on the semantics of definites and indefinites are examples of this view.

A "The understanding of sentences often requires that we construct contexts in which those sentences can be interpreted." Van der Sandt (1988)

B "To presuppose nothing means to place no requirements on the initial context." Heim (1982)

It is easy to read statements A and B in such a way that they say nothing controversial. A could be taken to be a straightforward explanation of truth-conditional meaning and B an unexceptional definition of presupposition. However, it is our contention that such readings of statements A and B are mistaken.

If, instead of A, van der Sandt had written A', then we could be sure that what he intended was an explanation of what it is to understand the propositional content of a sentence.

A' "the understanding of sentences *always* requires that we *can* construct contexts in which those sentences *could* be interpreted."

However, it is possible and we believe it is correct, that what van der Sandt (1988) actually means is that to obtain the intended information encoded in some token of a sentence a hearer often has to construct a context and then add the propositions contained in the sentence to it. We find this model of interpretation implausible. It is our contention that in general, in real discourses, hearers do not have to construct contexts for sentences using clues contained in those sentences. The utterances of discourses take place *in* contexts, and so hearers have contexts available with respect to which they can interpret utterances. However, since the propositional content of a discourse is not restricted to the physical situation in which it occurs, it is not always the case that the intended context for some utterance is the context of the discourse situation. Even so, it is not necessary to assume that these contexts are *constructed* by the hearer. An equally valid assumption is that alternative contexts are *identified* or *chosen* by him or her. While there *are* utterances which contain conventional discourse markers employed to constrain the kind of context which the hearer can choose, we claim that it would be a mistake to use these as templates for discourses in general. For example, any utterance which begins with one of the phrases in 2.20 is clearly marked as a case in which a new context must be chosen. These markers all indicate that the correct context is separate from the context in which the previous discourse has been grounded, and they each specify some features of the context, e.g. *Long ago and far away* and *once upon a time* indicate that the following statements are not concerned with the real world.

2.20. Once upon a time ...
A man walked into a bar ...
Long ago and far away ...
An Englishman, a Scotsman and an Irishman ...

To these atypical sentence/context pairs we can add the example sentences about John and Mary which are used in semantics textbooks and most unfortunately, the sentences about the King of France which appear in expositions of presuppositional theories.

Heim's (1982) definition of presuppositions which rests on equating a lack of presuppositions with the null context (statement B above) seems equally harmless. But this impression is the result of our interpreting Heim's statement as a description of the knowledge states of discourse participants. We could accept the claim that a speaker has no presuppositions if he or she places no restrictions on the hearer's choice of contexts, or that a hearer has no presuppositions if he or she places no restrictions on which contexts he or she will consider as potential contenders against which to interpret an utterance. However, since Heim assumes presuppositions to be features of *sentences*, we must interpret A as a claim that when a sentence has no presuppositions, the sentence places no restrictions on the features of its own context. And we must contrast this with the other possibility that a sentence which does have presuppositions does impose restrictions on the features of its context. Heim's claim, then, is open to the same objections which were made to Van der Sandt's: that it is contexts which constrain the utterances which occur in them, not sentences which constrain their contexts.

We must conclude that it is not possible to combine a sentence-level viewpoint with appeals to discourse level phenomena such as context and background knowledge. However, one sentence level approach to the problems surrounding failed existential presuppositions which we believe is particularly interesting and deserves close attention is J.D. Fodor's theory of reference (Fodor 1979). In the following two sections we will review Fodor's paper in some detail and discuss Burton-Robert's critique of her approach to presupposition failure.

2.3. Fodor's Possible Worlds Approach.

J.D. Fodor's 1979 paper *In defense of the truth-value gap* is worth discussing at some length since some of the analysis in the later chapters of this thesis will either build on the analysis presented by Fodor or re-analyse her data in a different way.

Before presenting her analysis, Fodor briefly considers and then rejects two alternative analyses for aspects of the problem of failed existential presuppositions which both deserve a closer look.

2.3.1. Irrelevance Analysis

Fodor contemplates a potential distinction between the results of the failure of relevant presuppositions and the failure of irrelevant presuppositions which could be employed to produce an explanation for the truth-value assignments to sentences 2.21 (*truth-valueless*) and 2.22 (*false*).

2.21. The capital of Antarctica is remarkably law-abiding.

2.22. We are standing in a street in the capital of Antarctica.

The analysis is that if a hearer of an utterance of 2.22 knows where the utterance is taking place and knows that this location is not in the capital of Antarctica they will be able to assign the truth-value *false* to it without any kind of access to the capital of Antarctica, and that the failure of the presupposition that the capital of Antarctica exists is therefore irrelevant to the truth-value assignment. In contrast, a hearer of an utterance of 2.21 must have some kind of access to the capital of Antarctica to determine whether or not it is law-abiding and since the entity is non-existent this access is not available. This analysis has the result that the presupposition that the capital of Antarctica exists is relevant, and the failure of this presupposition, brought about by the non-existence of the capital of Antarctica, prevents the assignment of a truth-value to 2.21.

Fodor rejects this analysis on the grounds that it is impossible to maintain any distinction between relevant and irrelevant presuppositions which depends on the notion that we require access to non-existent entities to assign truth-values in some contexts. She argues that such

sentences as 2.23 would have to be classified as containing instances of irrelevant presuppositions since a hearer of 2.23 could establish the identity of the speaker's home town and make sure that it was not the capital of Antarctica without any kind of access to the capital of Antarctica.

2.23. The capital of Antarctica is my home town.

Fodor contends that if we accept this, then we must accept that we should be able to make a list of all law-abiding places and, having checked that the capital of Antarctica is not on the list, be able to assign a truth-value of false to sentence 2.21, again without access to the capital of Antarctica. Since sentence 2.21 is perceived to be truth-valueless, Fodor concludes the divergent truth-values for 2.21. and 2.22. cannot be accounted for in terms of a distinction between relevant and irrelevant presuppositions.

2.3.2. Multiple World Reference.

The second analysis which Fodor considers, but ultimately rejects, rests on the claim that the capital of Antarctica although not existing in the real world, exists in an infinite number of other possible worlds in each of which it has some distinct level of law-abidingness. Our inability to assign a truth-value to 2.21 can be explained in terms of the multiple world reference account in either of two ways: either a sentence containing a reference to the capital of Antarctica is about the entire set of worlds in which the capital of Antarctica exists, or it is about one of the worlds whose identity we are unable determine. If the former, then it is both true and false of the set of worlds as a whole that the capital of Antarctica is law-abiding in those worlds, and therefore we cannot say that it is true and we cannot say that it is false. If the latter, we simply cannot say whether the particular member of the set of potential reference worlds is in the sub-set of worlds in which the capital of Antarctica is law-abiding or in its complement; and we are again unable to assign a truth value of either true or false.

2.3.3. Thin fictions.

Having rejected both analyses outlined in the last two sections, Fodor presents her proposals for the source of the truth-valuelessness of sentences with failed presuppositions. Her account rests on the assumption that fictional worlds such as those containing Winnie-the-Pooh, Sherlock Holmes and Mr Chekhov (the navigator of the USS Enterprise) are underspecified. By underspecification, Fodor means that it is not necessarily true of a fictional world that some proposition predicated of an entity in that world is either true or false. She does not discuss how it is decided which properties of a fictional world are specified and which are not but we can infer the rule from the examples she gives.

2.24. Winnie-the-Pooh likes/hates honey.

2.25. Winnie-the-Pooh likes/hates Mondays.

On the strength of these examples, we assume that, in Fodor's view, the only properties which are specified in a fictional world are those contained in propositions which are explicitly stated in the works of fiction by which the world is invoked.

Fodor then proposes that non-existent individuals be treated in the same way as fictional individuals, so that sentences containing references to non-existent entities should be regarded as "thin fictions" which refer to what she calls "drastically underspecified worlds" (Fodor, 1979: 204). For example, she argues that if Winnie-the-Pooh exists in an incomplete world in which the truth-value of the proposition that he has the property of liking honey is specified, but it is not specified whether or not he likes Mondays, we can account for the lack of a truth-value in sentence 2.21 by arguing that the capital of Antarctica exists in a world which is unspecified for almost everything. In this world, Fodor argues, the only specified properties are those which must be true of any capital of Antarctica, given that the phrase **the capital of Antarctica**, is used in a literal and conventional sense; for example, the following sentences may be assigned truth-values.

2.26. The capital of Antarctica is a place.

2.27. The capital of Antarctica is not in Africa.

We have strong reservations about Fodor's notion of thin fictions and will show in chapter 4 section 4.4.2. that it is unnecessary. However, other aspects of Fodor's approach are extremely interesting and well-motivated, particularly her approach to the question of cross-world relations which we will now review.

2.3.4 Cross-world relations

Fodor proposes that sentences can only be used to make statements about worlds in which their constituent referring expressions have referents. Besides sentences about the real, fully specified world, which necessarily have truth-values, and sentences about non-real underspecified worlds which might be truth-valueless, there are further sets of sentences which are about combinations of the different worlds in which their individual constituent referring expressions each have referents. For example, 2.28 is analysed as a statement about a pair of worlds, 2.29 about a triple of worlds, and 2.30 about a quadruple of worlds.

2.28. Lieutenant Uhuru bit Elizabeth Bennet.

2.29. Lieutenant Uhuru introduced Elizabeth Bennet to Agent Scully.

2.30. Lieutenant Uhuru prefers Elizabeth Bennet to either Agent Scully or Melanie Wilkes.

To account for the pattern of truth-values perceived for such sentences as 2.28 - 2.30 Fodor divides verbal predicates into three types: affectless, symmetric, and asymmetric predicates, and distinguishes between two types of relation which can be described by verbal predicates: same-world and cross-world relations. An affectless predicate is one whose argument or arguments pick out referents which are not affected by the occurrence of the event the verbal predicate describes: for example, **resemble** and **be taller than** are affectless predicates. One

feature of these predicates is that they may describe cross-world relations. That is, they may be combined with two arguments whose referents exist in different worlds from one another to produce sentences which are potentially true. While other types of predicates can combine referents across worlds, Fodor argues that only affectless predicates can combine with a pair of arguments in either order to produce sentences which are potentially true. For example:

2.31. Oliver Sacks resembles Santa Claus.

2.32. Santa Claus resembles Oliver Sacks.

Asymmetric affect is a feature of two-place verbal predicates which have one argument whose referent is affected in some way by the verb event and one whose referent is not; for example, **remember** (subject affect), **frighten** (object affect) and the predicates which are classified as opaque (subject affect). Fodor claims that asymmetric predicates can cross world boundaries if and only if the affected argument has a referent in the real world. For example,

2.33a. John Major remembers Desperate Dan.

2.33b. *Desperate Dan remembers John Major

2.34a. Tank Girl frightens John Major.

2.34b. *John Major frightens Tank Girl.

2.35a. John Major admires The Terminator.

2.35b. *The Terminator admires John Major

Symmetric affect predicates are those, both of whose arguments have referents which are affected by the verb event e.g. **kick** and **shoot**. These predicates cannot cross-world boundaries. For example,

2.36. *Bilbo Baggins kicked Mrs Thatcher.

2.37. *Mrs Thatcher shot Bilbo Baggins.

Fodor herself recognises that there is a range of exceptions to her general predictions, and she proposes a set of pragmatic principles for interpretation which can be invoked to account for

the apparently exceptional truth-value assignments observed. Fodor bases her analysis on the assumption that the real world has exceptional status in the set of possible worlds and she proposes a Real World Principle (RWP):

> "we tend to relate sentences to the real world if it is possible to do so ... [An expression, if] it does refer in the real world is understood as so referring."
>
> (Fodor, 1979: 217)

The RWP would predict a truth-value of *false* for both of the sentences in 2.38.

2.38a. The King of France danced with a Dutchwoman.

2.38b. A Dutchwoman danced with the King of France.

On the basis of her classification in terms of affect symmetry Fodor would predict that 2.38a and 2.38b are both false since **dance with** is a symmetric predicate, which cannot cross-world boundaries, and the Real World Principle requires the **a Dutchwoman** to have a referent in the real world, while there is no referent for **the King of France** in the real world. However, while 2.38b is intuitively false, 2.38a is perceived as truth-valueless; i.e. both arguments are taken to refer in a non-real under-specified world with respect to which the sentence might be either true or false. To account for this, Fodor, proposes a second pragmatic principle; the Principle of Left Association (LAP):

> "An indefinite noun phrase is taken to refer to an individual ... in the same world as the noun phrase on its left, if there is one, and otherwise to an individual in the real world."
>
> (Fodor, 1979: 218)

This ensures that **a Dutchwoman** in 2.38a refers in the same world as **the King of France**. However, the Principle of Left Association makes a wrong prediction about such sentences as 2.39a and 2.39b, both of which are intuitively false.

2.39a. A friend of mine shook hands with the King of France.

2.39b. The King of France shook hands with a friend of mine.

To account for these sentences Fodor suggests yet another competing pragmatic principle, The Principle of Maximising Truth-values (MTP), which stipulates that **a friend of mine** must refer in the real world if this allows us assign a truth-value to either 2.39a or 2.39b: "we tend to evaluate a sentence with respect to a world in which it has a truth-value." (Fodor, 1979: 217). In this case, the RWP and the MTP predict the same referent for **a friend of mine**, i.e. an entity in the real world, and it is plausible that the two principles combined outweigh the LAP. However, the same should be the case in relation to sentence 2.38a. Fodor cannot explain why the RWP and the MTP do not outweigh the LAP in this sentence. This is all the more puzzling since in sentences where the RWP is not applicable, e.g. sentence 2.40, the MTP *alone* outweighs the LAP, with the result that the sentence is false.

2.40. Bilbo Baggins toured the Enterprise on an open day.

If Fodor is to salvage her pragmatic principles, she must provide an explanation for the failure of the MTP in 2.38a. She concludes that there is some division between those entities which we will allow to jump worlds (a Dutchwoman) and those which we will always interpret with respect to the real world (a friend of mine). She acknowledges that this division is fuzzy and does not offer any criteria for determining the ability or non-ability of some entity to jump worlds but offers the following observation

> "the real world entities which we are prepared to acknowledge as existing in non-real worlds seem to be just those publicly familiar entities which we are not surprised to find referred to in fictions".
>
> (Fodor, 1979: p219)

This statement provides not even the beginning of an adequate explanation. It is certainly unarguable, but we are convinced that this is because it is a tautology. If referring expressions in fictions pick out entities in other worlds then of course the set of other-worldly entities coincides with the set of fictional referents. Without some means of dividing entities which can shift worlds from entities which cannot, in a non-circular way, the explanatory

power of the three competing pragmatic principles is extremely dubious; as they stand it is necessary to know the desired truth-value in order to predict the weighting of the principles which will produce it[8]. Although we ultimately reject Fodor's account it is worthwhile reviewing a critique of her argument which contains some interesting observations that will prove useful in the construction of the current theory.

2.4. Burton-Roberts' Critique of Fodor

Burton-Roberts (1989) contains a sustained and severe criticism of various aspects of Fodor's account of which three points are relevant to the current thesis: his defence of the concept of irrelevant presuppositions, his attack on underspecification, and his discussion of choice among non-real potential reference worlds.

2.4.1. Irrelevant Presuppositions.

Burton-Roberts supports the view that failed existential presuppositions which do not result in truth-valuelessness in the presupposing sentence can *always* be identified as irrelevant presuppositions; an analysis which Fodor rejected. Recall that Fodor suggested that the failed existential presuppositions in sentences such as 2.41 and 2.42 were equally irrelevant.

2.41. The King of France is bald.

2.42. I am standing next to the King of France.

Fodor denies that the procedure of checking the identity of the individuals standing next to the speaker of 2.42 is qualitatively different from the procedure of checking the identity of the members of the set of bald individuals. Burton-Roberts, however, argues that there *is* a qualitative distinction between the two procedures. He claims that it is possible to know that

---

[8]In chapter 6, we will show that these principles are unnecessary since, the combination of the existential ranking of linguistic expressions proposed in chapter 3, with the structure of the set of possible worlds proposed in chapter 4, allows us to predict intuitive truth-values in a coherent way.

the list of individuals standing next to the speaker of 2.42 is complete, without knowing whether or not there is a French King but denies that "it is possible to know that the set of bald individuals does not contain the King of France without knowing that there is no French King." (Burton-Robert, 1989: p.200). Burton-Roberts gives as a further example of an irrelevant failed presupposition, the following sentence:

2.43. Max spent the day at the local swimming pool. (20 in Burton-Roberts, 1989: p201)

He argues that whether or not we know whether there is a referent for **the local swimming pool** we can assign a truth-value to this sentence if we know what activity Max undertook on the relevant day. This is indisputable, but we believe that sentence 2.43 can also be, in fact must be, interpreted as false by some hearer who knows Max and knows that there is no local swimming pool. Burton-Roberts would predict truth-valuelessness for sentence 2.43 in this context. On the other hand, he would predict that the non-existence of the King of France is irrelevant to the truth-value assignment for 2.44.

2.44. In 1978, the King of France shook hands with a certain labour MP.

He would argue that the assignment is made possible by our ability to make a complete list of the individuals who shook hands with labour MPs during 1978 and would predict that a hearer does not have to know that the King of France does not exist to be able to assign the correct truth-value, *false*, to sentence 2.44.

Fodor, in contrast, would propose a combination of the RWP and the MTP outweighing the LAP in the interpretation of sentence 2.44, to predict the truth value which speakers assign to the sentence, i.e. *false*. However, she could not explain why the weighting of the three principles is different with respect to 2.38a which is generally perceived to be truth-valueless.

We find both of these arguments extremely implausible. With respect to Burton-Roberts approach, it seems that an analysis of 2.44 as a false statement crucially depends on the

knowledge that the King of France does not exist. A hearer must interpret the expression *a certain Labour MP* as a reference to a real world entity, must interpret *the King of France* as a reference to an entity which does not exist in the real world and must interpret *shook hands with* literally, as a description of a physical action, in order to interpret an utterance of the sentence as false. We predict that only a hearer who did not know that the King of France does not exist, and who accepted *the King of France* as a reference to a real world entity would be unable to assign a truth-value to this sentence. Burton-Robert's discussion of the analysis in terms of irrelevant presuppositions provides a further incentive to account for the truth-value of 2.44 along the lines of Fodor's account, but without any reliance on ad hoc principles.

2.4.2. Underspecification

Burton-Roberts also takes issue with Fodor's proposals concerning thin fictions which rests on an assumption of the incompleteness of non-real worlds. He claims: "Fodor's account exhibits a confusion between an incompletely specified world and an incomplete world." (Burton-Roberts, 1989: p.212) and argues that while non-real worlds may be incompletely specified, no world can actually be incomplete. We believe that the assumption that Fodor is confusing two separate issues is mistaken and that the notion of an incomplete world is *precisely* what she intends to underpin the notion of thin fictions[9].

Having discounted the possibility of truly incomplete worlds, Burton-Roberts argues that Fodor's use of the concept of thin fictions can only be understood as a claim that a sentence such as 2.41 *has* a truth-value in the world where the definite description has a referent, but that *no* truth-value can be assigned to it; an argument which he reject completely. He appears to find the notion that a sentence could have a truth-value which no one can assign to it utterly unacceptable and presents this as evidence of a flaw in Fodor's argument.

While we do not believe that Fodor actually intended to imply that there was a distinction

---

[9] Like Burton-Roberts, we find the notion of worlds which have the property of being incomplete difficult to accept. We deal with the possibility of underspecification at some length in chapter 5 (section 5.3.1.) during the discussion of the nature of possible worlds..

between a sentence's having a truth-value and the possibility of that truth-value's not being assigned to it, we cannot agree with Burton-Roberts that the notion is unacceptable. In fact we will show in chapter 4 section 4.3.1. that this is an extremely useful and important distinction.

2.4.3. Choice of reference world.

Burton-Roberts (1989) argues that Fodor's account does not predict that sentence 2.45 is perceived as truth-valueless but that we can deduce from her predictions that it is both true and false.

2.45. The King of France is bald.

He claims that Fodor's account predicts that 2.45 is false. He draws attention to her claim that sentences are about worlds in which their referring expressions have referents, which would ensure that 2.45 cannot be about the real world but he argues that this claim is contradicted by her claim that it is possible to make an exhaustive list of every bald individual and check to see if the King of France is on it. *This* claim, Burton-Roberts contends, has the result that 2.45 can be taken to be a false statement about the real world. Burton-Roberts then argues that Fodor's analysis allows us to interpret 2.45 as true.

> "if it is clear from the presence of **the King of France** (in 2.45) that the speaker is speaking about a set of worlds containing the King of France it is equally clear from the presence of **is bald** which more specific, more delimited set within that set he is talking about ... he is and must be talking about that set of worlds in which there is a bald King of France ... having allowed speakers a choice as to the worlds of which they speak we cannot but allow that what they say of those worlds is vacuously true."
>
> Burton-Roberts (1989: p.214)

We agree to a certain extent with Burton-Roberts objection, in that if there is a *free* choice among prospective reference worlds then truth-value assignments can never be predicted, although we do not agree that Fodor's theory either advocated or allows such a free choice.

We do agree, however, that Fodor's insistence that sentences are about *worlds*, leads to a contradictory prediction that sentence 2.45 can and cannot be used to make a statement about the real world. The analysis presented in chapter 6 section 6.2.5 includes a coherent means to constrain the choice of reference world for this sentence to ensure that it is perceived as a truth-valueless statement about an entity in a non-real world.

2.5. Beyond presuppositions.

The discussion of competing presuppositional analyses in the previous sections was intended to do three things: first, to bring into focus the area to be covered, second, to highlight the valuable insights to be found in what we consider to be the best treatments, and third, to identify the problem areas which no current theory of presupposition can handle adequately. What the discussion showed is that neither semantic nor pragmatic analyses of presuppositions can account for the pattern of truth-values observed. It clearly does not appear to be the case that a simple correlation between presupposition failure and truth-valuelessness is tenable. Indeed, existential presuppositions, whether upheld, failed or cancelled, do not appear to provide a rich enough framework in which to discuss the range of references to real, fictional and non-existent entities, and the related truth-values of positive and negated sentences which contain them. We will now consider an approach in which the category of presuppositions is abandoned altogether.

In a paper titled *Accommodation and the myth of presupposition: The view from discourse*, Paul Werth argues that a logical impasse results if a theory which attempts to deal with the problems we are currently addressing is restricted in its perspective to the isolated sentence, but at the same time invokes the notion of the context (Werth 1993).

> "Since 'the context', however defined, is clearly not part of the sentence, but belongs to the level which contains the sentence (i.e. discourse) the problem ought to be treated as one of discourse, i.e. at the minimal level of inclusion for all the elements involved."
>
> (Werth, 1993: p.46)

He goes on to make the controversial claim that presuppositions are the "epiphenomena of an inappropriate perspective" and argues that, from the discourse viewpoint, there is no need for the linguistic category of presupposition in any form.

Central to Werth's claim that there is no such thing as presupposition is the idea that the range of semantic phenomena which has been lumped together under that heading in the past are of two different types.

> "...accommodation, the introduction of new information non-assertively, is not merely a funny kind of presupposition, but is in fact functionally totally different."
>
> (Werth ,1993: p39)

This theory assumes that in any discourse there is a common ground (CG), which includes the set of propositions previously expressed in the discourse, the propositions entailed by them and the propositions they activate from the set of potentially relevant propositions in the knowledge bases of the speaker and the hearer. Werth proposes that we can categorise the propositions contained in some utterance in the following way: either as propositions already present in the CG which represent backgrounded information (BI) or as propositions not in the CG, which represent new information being considered for inclusion in the CG. Werth argues that most of what are conventionally called presuppositions can be seen as references to backgrounded information and that the rest, conventionally analysed either as failed presuppositions causing truth-valuelessness or as presuppositions requiring an accommodation mechanism, can be seen as a special means of presenting new information which he calls unconventional assertions (UA). Werth also claims that information introduced by an unconventional assertion will typically be unimportant or uncontroversial. For example in sentence 2.46, the six bracketed clauses represent propositions encoded in conventional assertions

2.46. [I have a neighbour] and [he was a keen gardener]. [I also have cats] and [they ruined my neighbour's garden]. [My neighbour took up sailing] and [this was because of my cats].

The same propositions, in sentence 2.47, are presented as backgrounded information. If 2.47 is assumed to follow 2.46 in a discourse the presentation of the propositions in this way reflects the fact that they are already part of the CG. If 2.47 is the initial utterance in a discourse which takes place between strangers the propositions must be analysed as new pieces of information being presented in unconventional assertions.[10]

2.47. I am slightly embarrassed that my previously garden-mad neighbour has taken up sailing because of my cats' destructive activities.

Werth claims that although UAs are primarily employed to present uncontroversial new information this feature of their use is pragmatic, therefore cancellable and exploitable i.e. a speaker may flout the rule to present a startling or otherwise very informative proposition using a UA, to give the impression to the hearer that they *consider* the proposition unsurprising (or would like the hearer to believe that they do). For example, 2.48a and 2.48b present a (typically) very informative proposition as respectively a conventional and an unconventional assertion[11].

2.48a. [I made a million] and moved to the Costa.

2.48b. After [I made my million] I moved to the Costa.

Werth identifies a range of means by which a proposition can enter the common ground of a discourse: these include via conventional assertion or unconventional assertion, by being an entailment or an implicature of some other proposition in the common ground or by being part of the general world knowledge of the discourse participants. In contrast to the constraints on unconventional assertion, Werth argues that a proposition can be referred to by means of a so-

---

[10]The distinction between unconventional assertions and references to backgrounded information is similar to an idea in Wilson and Sperber (1979) in which entailments and presuppositions are collapsed into an ordered set of truth conditions of one formal type. Wilson and Sperber argue that a speaker can pragmatically indicate the most important entailments of an utterance by linguistic means (either structural or intonational). They describe the choice between uttering two conjoined main clauses to indicate two propositions of equal importance or uttering a main clause/subordinate clause structure to indicate that one proposition is less important than the other because it contains information that is either old or uncontroversial.

[11]In section 3.2.2.6 of chapter 3 we discuss in detail question of whether there are limits on the range of existential propositions which can be presented unconventionally.

called presupposition triggering expression regardless of how it originally entered the common ground of the discourse. For example, any of the propositions presented as unconventional assertions in 2.47 (read as a discourse initial utterance) can be treated as backgrounded information. If the speaker of 2.47 were to continue by uttering sentence 2.49 the pronouns **they** and **it,** would depend for their interpretation on the existential propositions in 2.49a.

2.49. They made a real mess of it.

2.49a. My neighbour has a garden.
I have cats.

This a controversial claim and it is dealt with in detail in section 3.2 of the following chapter when we consider the existential contribution of a range of linguistic expressions individually. For the moment let us accept that propositions which have been *offered for inclusion* in the common ground by any means can be referred to as backgrounded information.

Werth's analysis is very interesting but since his intention is solely to provide a basis for the analysis of discourses and texts he does not address the interesting questions of what his analysis predicts for the classical puzzles of ambiguous negation or references to non-existent entities. We now discuss the light his approach might throw on the puzzles in these areas.

2.5.1. Negation Without Presuppositions.

Werth's analysis results in an apparent ambiguity, in isolated sentences, between an expression being interpreted as an unconventional assertion of new information, or as a reference to old information. We will describe this as an ambiguity between anaphoric and assertive readings of the expression. If we can show that only existential propositions associated with anaphoric expressions can be negated, then we will have solved the problem of the source of the ambiguity or vagueness of negation scope. In fact, we will be able to show that negation is not ambiguous, but that when sentences are evaluated out of context it is not always clear which proposition in a sentence is being negated.

Clearly, negating some proposition being asserted by the use of some expression would constitute a contradiction. To illustrate this let us consider an utterance of 2.50 in a context where the proposition in 2.50a is not part of the previously established common ground. In such a context, the use of the expression **my nephew Alastair** constitutes a unconventional assertion of the proposition that the speaker has a nephew called Alastair.

2.50. My nephew Alastair is freckly.

2.50a. I have a nephew called Alastair.

If in this context, the speaker were instead to produce an utterance of 2.51 the utterance could contain both an assertion and a denial of the single proposition that the speaker has a nephew called Alastair.

2.51. My nephew Alaistair is not freckly.

2.51 in this context, and on this reading, therefore would be a a contradiction. On the other hand, in a different context, in which the proposition in 2.50a *is* part of the established common ground shared by speaker and hearer, i.e. if both discourse participants know that the speaker of 2.51 has a nephew called Alastair, then 2.51 contains no contradiction since only the property of freckliness is being denied, not the existence of the individual named Alastair.

A more interesting situation can arise, again when 2.50a is not part of the common ground and the speaker of 2.51 does **not** have a nephew called Alastair. Suppose that in such a context, the other discourse participant utters 2.52 in which 2.50a is presented via an unconventional assertion.

2.52. Your nephew Alastair's freckly, isn't he?

A subsequent utterance of 2.51, in *this* context contains a reference to a proposition already asserted, which *can* therefore be negated by the speaker without contradiction.

The two states Werth outlines for propositions (the state of being newly asserted and the state of being established) can be seen as two extremes on a scale. Current theories of discourse suggest that new propositions spend a period of time during which they are under negotiation for inclusion in the subsequent common ground. (Brown and Yule, 1983). Before this, propositions must be offered for inclusion by a speaker and this may take place by various means: conventional assertion, unconventional assertion, implicature and references to unused propositions of background knowledge.

It is useful to view the status of new propositions in terms of a scale of commitment, with conventional assertions being associated with the highest level of commitment from the speaker and general background knowledge being associated with the lowest level of commitment from both speaker and hearer (Prince, 1981).

We propose that different levels of commitment can be seen as differences in the degree of responsibility discourse participants must accept for the inclusion of propositions in the common ground. If a speaker explicitly offers a particular proposition via a conventional assertion he or she must accept a high degree of responsibility for its subsequent presence in the common ground and cannot cancel it without contradiction. For example a speaker of 2.53 has a high degree of responsibility for the inclusion in the discourse of 2.54 and it would be contradictory for him or her subsequently to utter 2.55.

2.53. An apprentice organ-builder from York has invented a time machine.

2.54. There is an apprentice organ-builder from York.

2.55. There are no apprentice organ-builders these days.

If the hearer does not reject the proposition in 2.53 during the initial negotiating period, then he or she must subsequently accept some degree of responsibility for its presence in the common ground, but the hearer is able to cancel the proposition in either of two ways: they might reject the proposition outright when it is offered for inclusion, or attempt to remove the proposition from the background if additional later information shows it to be inconsistent with other propositions in the context. Neither of these two methods of cancellation involve

the hearer in any measure of contradiction. (Brown and Yule, 1983; Werth 1993)

At the other end of the scale, if a proposition is part of background general knowledge then we claim that both the speaker and hearer have a much smaller amount of responsibility for its presence and either one can cancel the proposition without contradiction. The ordering is interesting, with respect to negation, because as the discourse participant's level of commitment to a proposition decreases, any negation of the proposition appears less like denial negation and more like descriptive negation (in the sense of Kempson 1977). For example, we can compare 2.56, where A's utterance contains a conventional assertion of the proposition negated by B's utterance, with 2.57, where there is very little possibility of identifying the source of the proposition 2.57a which is being negated.

2.56. Speaker A: The King of France is bald.
Speaker B: The King of France is not bald.

2.57. There is no God.
2.57a. There is a God.

The result is that the negation in 2.57 seems less like a denial than the negation in 2.56. Nevertheless, we claim that 2.57 *does* contain an instance of a denial of a proposition in the common ground which can be evaluated as either true or false by hearers. In contrast, 2.58 which has negative structure does not constitute an instance of negation in the current context since the proposition in 2.58a is not in the common ground.

2.58. There is no Flast.
2.58a. There is a Flast.

If we make this assumption that speaker and hearer are jointly responsible for negotiating the context we can explain why the example given by Kempson (1977) from Strawson (1950) and repeated below is not contradictory.

> "Neither Aristotelian nor Russellian rules gives the exact logic of any expression of ordinary language: for ordinary language has no exact logic."
>
> (Kempson, 1977: 152)

Contrary to Kempson's claim, the proposition that there is an exact logic for natural language is not being asserted in this sentence. The proposition is already present, under negotiation in the common ground. But even though the sentence appears in a written monologue this does not mean that somewhere in the preceding text, a statement of Strawson's must have asserted, entailed or implied this proposition. He may assume that it is part of the world knowledge of his readers, or more interestingly, he may assume that the reader will have added the proposition to the context because they *mistakenly* perceive it to be entailed or implied by his preceding argument.

We find the initial analysis of negation without presupposition to be encouraging enough to motivate an attempt to re-evaluate the problems of references to non-existent entities without using the category of presupposition. In the following chapter we will examine closely the validity of Werth's distinction between unconventional assertions and references to backgrounded information when it is applied to existential propositions introduced by means of different types of nominal expressions.

Chapter 3
Existential Features of Linguistic Expressions

In the introductory chapter, we made use of the idea that there is a plurality of existential states when we chose to describe some entities as "existentially questionable" rather than simply non-existent. This might be taken to imply that there are further existential possibilities besides some entity's existence or its non-existence. It should not be imagined that we assume either that there is general acceptance of there being such a plurality or, even if there were, that the nature or number of possible existential states is established. On the contrary, we acknowledge the strength of the intuition that existence must, in some basic sense, be a binary property. However, there is a crucial distinction to be drawn between existential properties expressed linguistically and objective existential properties; that is, between existential aspects of the meaning of nominal expressions and the actual existence or non-existence of the objects to which these expressions refer. It is existence as a property of objects which our intuitions dictate must be binary[1]. There is no compelling motivation, in a theory which embraces pragmatic aspects of meaning, to view linguistic existential properties as intrinsically binary. In this chapter, we will propose an approach to the existential aspects of the semantics of linguistic expressions which is scalar rather than binary. First we will define our use of the term existential force (EF) and discuss its location, that is, identify what kind of expressions are the bearers of EF.

3.1. Existential Force.

3.1.1. The definition of EF.

Existential features of the meaning of linguistic expressions have been given a number of different names in the semantic and logical literature; for example, existential force, existential import and existential commitment (Kempson, 1977; McCawley, 1980; Burton-Roberts, 1989). However, the underlying analyses are essentially the same. McCawley (1980), for

[1]In fact, as we discuss in chapter 4, section 4.1, if we recognise a variety of existential locations the result is that there appears to be a wider range of possibilities than this.

example, defines existential commitment as the commitment of a speaker of some sentence which contains a quantified NP to the proposition that the domain of that quantifier is not empty (McCawley, 1980: 112). In the terms of Hawkins' (1978) account of reference and (in)definiteness[2], where the contribution of definites and indefinites is also seen from the point of view of the speaker, existential force could be defined as a signal of the commitment on the part of a speaker to the existence of the object to which some NP refers. In the theory proposed here, which views interpretation and truth-value assignment from the point of view of the hearer, the level of EF is defined as the level of acceptance, by a hearer, of the existence of a referent for some NP.

3.1.2. The Location of EF.

Formal accounts of the semantics of NPs tend to view EF as a property of the class of logical quantifiers which represent, at the level of logical form, both definite and indefinite determiners as well as the set of natural quantifiers. In Montague Grammar, (Dowty, Wall and Peters 1981; Cann 1993) and Generalised Quantifier Theory (Barwise and Cooper, 1984; van Benthem, 1986; Partee, ter Meulen and Wall, 1990) **the**, **some**, **sm**, **both**, **many** and **a few** are translated into logical form as quantifiers with a positive EF value with the result that the use of **the** in an utterance of sentence 3.01 commits a speaker to the existence of a leprechaun; the use of **both** in sentence 3.02, to the existence of two leprechauns, and the use of **many** in 3.03, to the existence of more than two leprechauns.

3.01 The leprechaun hates vodka.

3.02 Both leprechauns hate vodka.

3.03 Many leprechauns hate vodka.

In contrast, **every**, **any**, **few** and **no** are translated as quantifiers with a negative value for EF so that utterances of sentences 3.04 - 3.07 do not commit their speakers to the existence of leprechauns.

---

[2] See chapter 3 for a review of Hawkins' approach

3.04 *Every leprechaun hates vodka.*

3.05 Any leprechaun hates vodka.

3.06 Few leprechauns hate vodka.

3.07 No leprechauns hate vodka.

There is a potential problem, however, with the association of EF with quantifiers and determiners; if EF is located in the determiner then any expression with either positive or negative EF must be analysed as containing a determiner. We will now discuss the merits of positing determiners for the remaining types of nominals i.e. names, pronouns, mass terms and bare plurals.

It is now uncontroversial for syntactic theories to assume that pronouns are determiners. This analysis is proposed in Hudson (1984) and the case is argued persuasively in Abney (1987). From a semantic point of view also, an analysis which assumes that pronouns are determiners is well-motivated. The pronoun in 3.08, for example, can be analysed as a combination of an inflected determiner which encodes case, number and gender and a null descriptive predicate whose content is recoverable either from the preceding discourse in the case of anaphoric uses, or from the physical situation of utterance in the case of deictic uses[3].

3.08. He finished the vodka.

This analysis brings the distribution of semantic content in pronominal NPs into line with the distribution of semantic content in NPs with overt determiners **and** descriptive predicates.

The problem with bare plurals and mass terms is of a different kind. Here the descriptive predicate is phonologically realised but there is no overt determiner. However, these are NPs

[3]The deictic uses of third person pronouns are controversial however. It is not generally assumed that **he** and **she** can are deictic in the same way as **me**, **you**, **here**, **there** etc. (Lyons, 1975, 1977). It may be more accurate to propose that third person pronouns are only indirectly deictic; that they are linked with some pragmatic discourse feature such as an intonation pulse, a pointing gesture or an eye-flash; and that it is these gestures which have the deictic function.

for which we wish to propose positive EF. For example, the subject NPs in the following group of sentences all appear to be associated with positive EF.

3.09. Leprechauns hate vodka.

3.10. Leprechauns finished my vodka.

3.11. Vodka did for the leprechauns.

There are two possible solutions to the problem of the location of the EF in bare plurals and mass terms. We may decide either to view EF as a feature of referring expressions or to analyse the NPs in sentences 3.09 - 3.11 as having phonologically null determiners.

There is some independent motivation for positing null determiners for bare plurals and mass terms. If we consider the pairs of sentences in 3.12 and 3.13 we can see that bare plurals and mass terms are ambiguous between generic readings (in the a versions) and indefinite readings (in the b versions).

3.12a. Gin can cause tearfulness.

3.12b. Gin was the cause of my tearfulness last night.

3.13a. Voters are easily frightened.

3.13b. Voters were frightened by Norman.

The most straightforward way to deal with the ambiguous referential potential of these expressions is to set up two separate unambiguous phonologically null determiners which we can call **SM** and **GEN**. Clearly, **SM** which is simply a null variant of **sm** is to be analysed as having the same positive EF value; we discuss the EF of **GEN** in section 3.2.1.3..

Assuming that bare plurals and mass terms are amenable to an analysis which employs null determiners, the simplest and most cohesive analysis would be one in which we could handle names, e.g. 3.14 in the same way.

3.14. Bill Clinton finished my vodka.

In Montague Grammar, the semantics of names is handled by employing a process of type-raising to align them with the type of quantified NPs. This analysis is motivated by the requirement, which is a central concern of Montague Grammar, that expressions with identical syntactic distribution should be assigned to the same syntactic category and that entire syntactic categories should be also exhibit semantic uniformity: i.e. if NP is a real syntactic category and it includes quantified NPs, definites and names then all of these should be assigned to the same semantic type. There is, therefore, no difficulty in proposing null determiners for names since they must necessarily have determiner nodes at the level of syntactic representation (Cann, 1993). In purely syntactic treatments, also, a (usually empty) determiner node is now regularly associated with proper names (Hudson, 1984; Abney, 1987).

There are two types of empirical evidence in favour of a determiner node in the representation of proper names. The first of these concerns certain uses of proper names in English which allow the occurrence of overt determiners. For example:

3.15. The Mary Smith you were at school with, sounds just like the Mary Smith I was at school with.

This evidence alone is not entirely compelling, since it is possible to argue that **Mary Smith** in 3.15 is not being used as a name but that an utterance 3.15 only occurs if there has been some previous mention of Mary Smith and **Mary Smith** is being mentioned, or quoted, rather than *used* in an utterance of 3.15 and so its occurrence with a determiner is of no more theoretical significance than the co-occurrence of a determiner **the** with a conditional clause in sentence 3.16b.

3.16a. I'll be finished by Christmas, if all goes well.

3.16b. The if all goes well is the problem.

However, sentence 3.17 (which was uttered in a real discourse situation) provides an example of an occurrence of an overt determiner with a first mention proper name, **Janice Wilson**.

3.17. The Mandy McBain you were at school with, sounds just like the Janice Wilson I was at school with.

Even more compelling is the possibility of the co-occurrence of **a** with a name in such sentences as 3.18.

3.18. If there is a Duane Dibley in the hotel, would he please go to the foyer.

However, we could argue that **a Duane Dibley** is an elliptical form of **a person called Duane Dibley**, in which case, this is no more evidence of the co-occurrence of determiners with names than 3.19 is of the occurrence of strings consisting of determiners, adjectives prepositions and names.

3.19. The blue envelopes are for Manchester and **the red for London**.

On the other hand, if we read **a Duane Dibley** as meaning a person called Duane Dibley and **the Janice Wilson I was at school with** as the meaning the person called Janice Wilson whom the speaker was at school with, we could argue that, for example, any occurrence of the name Duane Dibley without an overt determiner could be analysed as meaning the person who has the property of being called Duane Dibley. Indeed, in such systems as that proposed by Keenan and Falz (1985), in which there are no expressions which directly denote individuals, the proper name **Duane Dibley** *is* analysed as denoting a set of properties, of which the property of being called Duane Dibley must be a member. We can conclude that, if we can accept the uses above as examples of naming, we must posit a determiner node in the syntactic representations of proper names, since this node is not always empty.

A stronger argument for this analysis can be made by considering cross-linguistic evidence. Many languages show the regular co-occurrence of overt determiners and names e.g. Italian

Gianni or *Il Gianni*. In such cases, the variants with and without overt determiners appear to be optional paraphrases without any of the semantic effects or restrictions associated with co-occurrences of names and overt determiners in English (L. Serratrice, pers. com.).

In the end, we must acknowledge that the existence of null determiners for proper names remains controversial. However, since names tend to present problems for all semantic and syntactic theories and since these problems are of many different kinds, we are not inclined to give too much consideration to the fact that names to not fit neatly into our view of NP semantic structure. Fortunately, the questionable status of null determiners for names will not present any practical problems for our theory of EF. We do agree that EF originates in the determiner of an NP. However, we can also think of NPs (or discourse referents) as having EF which they inherit from their determiner and, given two assumptions, this will allow us a straightforward treatment of the EF of names. The first assumption is that a determiner's EF is fixed; that is, no determiner can have positive EF when combined with one descriptive predicate and negative EF when combined with another. The second assumption, which we will defend in the section 3.2.2.4., is that there is no ambiguity in the EF of a name. In contrast to bare plurals and mass terms, if there is a normally empty determiner node associated with names, there is no evidence from the existential aspects of the meaning of names that more than one null determiner can fill it. We do not need access to the null determiner which occurs with a name in order to identify it, since we know that it is always the same one, therefore there would not be any serious negative implications for the current theory if it turns out to be impossible to defend the claim that the determiner node is actually there.

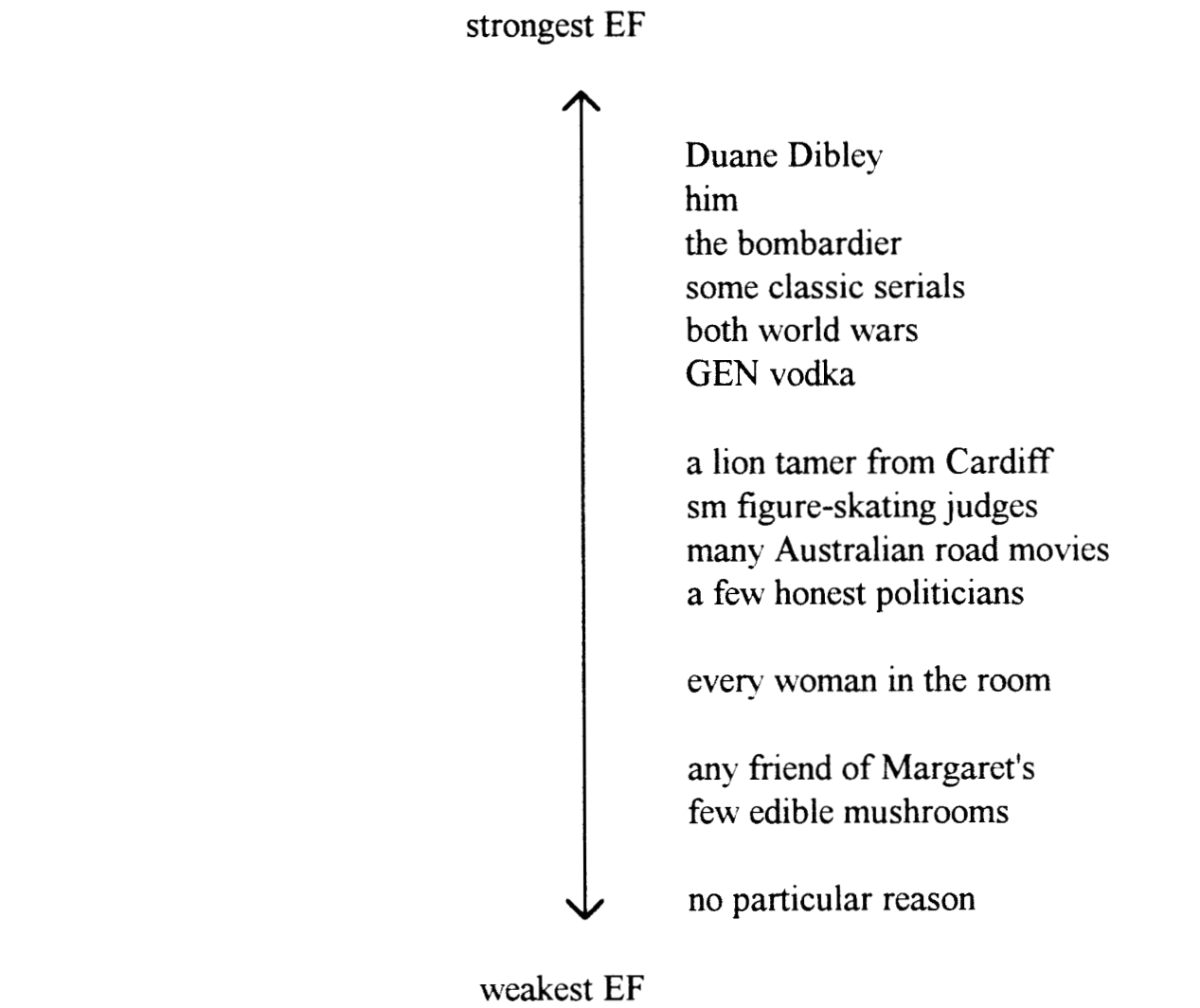
3.1.3. Binary or Scalar EF.

In Montague Grammar (Dowty, Wall and Peters 1981; Cann 1993) and Generalised Quantifier Theory (Barwise and Cooper 1984) **the**, **some**, **both**, **many** and **a few** are translated into logical form as existential quantifiers i.e. quantifiers with positive EF, while **every**, **any**, **few** and **no** are translated as quantifiers with negative EF. In fact, in all formal semantic theories the existential force of natural quantifiers and other determiners is

necessarily defined as a binary logical feature. However, we are convinced that a more accurate analysis of determiners is one which treats their existential properties as scalar. Table 3.1. shows the arrangement we propose for the scale of EF.

Table 3.1.

Scale of Existential Force



Despite the deviation from the traditional, formal view of EF as a binary feature which results from our placing the determiners on a scale, we do agree to some extent with the stance taken in formal accounts, and our arrangement does reflect the notion that it is possible to divide the determiners into two discrete sub-sets; those which are existential and those which are non-existential. The top two slots on the scale contain those determiners which are traditionally analysed as existential; **the, some**, **both**, **sm**, **many**, **a few**. The next lower two slots, which we will argue are associated with two weaker levels of EF, contain those determiners, **every**, **any** and **few**, which, if existential force is viewed as a binary feature, are lumped together with what we take to be the only non-existential determiner **no**. Our proposed scale is very similar to the binary analysis in that we retain the groupings at the top and bottom of the scale and the introduction of the two intermediate slots could be seen as no more than a blurring of the boundary. Since the abandonment of the binary distinction results in a system less formally elegant, and for which inferences are much more difficult to determine, we must face the question of whether we lose more than we gain by recasting EF as a scalar property.

Since our interests lie squarely in the area of the behaviour and use of referring expressions (and hence determiners) in real discourse we are generally happy to let empirical accuracy overrule considerations of formal elegance. In this instance, we propose to show that a scale of EF allows us to predict the behaviour of referring expressions in a much more realistic way than a binary split. Evidence for this claim is presented in detail in sections 3.2.2.-3.2.2.8 of this chapter.

3.1.4. Combining Linguistic EF with Objective Existence Types.

The question of the number and nature of kinds of existence is presently unresolved. Let us adopt the position discussed in chapter 4 section 4.1 that there are seven ways for entities to exist. Then, given the five possible levels of EF distinguished on the scale given in Table 3.1., there should be a potential thirty-five existentially distinct combinations of referring expressions and referents. It is our claim that all thirty-five potential combinations are possible; that an expression at any level of EF can pick out an object with any kind of

existence; and that since no collocation has any special idiomatic effect, that the result of the combination of any particular EF with any type of existence is completely predictable and no co-occurrence is deviant or borderline. The thirty-five possible combinations are laid out in Table 3.2 . The remainder of this chapter comprises a detailed discussion of linguistic existential features of all of these combinations.

Table 3.2.
Possible combinations of linguistic EF with objective existence types

	actual	**past**	**future**[4]	**fictional**	**mythological**	**physically possible**	**logically possible**
strongest EF	the Pope	Oscar Wilde	OJ's book	Becky Sharpe	the Tooth fairy	the Pope's wife	The Sheriff of Mars
	a friend of mine	a Roman slave	a friend of my son's*	a client of Holmes'	one of Santa's elves	one of Elvis's sons	an invisible man
	every Labour MP	every English serf	every child born in 1998+	every Lost Boy	every nasty troll	every female cardinal	every bi-lingual hamster
	any tourist guide	any cold war spy	any son of Henry IX	any Hobbit	any dragon	any WWIII heroes	any Welsh desert
weakest EF	no signing chimps	no Round-heads	no wife of Charles III	no party at Pooh Corner	no sirens' songs	no complete semantic theory	no black snow

* where the speaker has no sons.
+ since this thesis is being written in 1995

[4]The problematic distinction in this table, we acknowledge, is the distinction between future existence and physically possible existence. The question of whether there is a plausible distinction between the two will be discussed in the following chapter.

3.2. Levels of EF.

3.2.1. Strongest EF: Pre-establishment.

The first controversial feature of the proposed scale of EF is that there is a distinct level of existential force *above* the minimum level associated with the existence of a referent. That is, we are proposing that the traditional definite/indefinite distinction which separates **the** and **a** is seen as a distinction in EF, albeit one at a higher level on the scale than what has been traditionally labelled the existential/non-existential distinction which separates **many** and **any**. Our discussion of the contrasts at this end of the scale relies heavily on the definite and indefinite usage types classified by Hawkins (1978, 1984).

Hawkins employs two contrasts which he claims can distinguish among the range of definite and indefinite uses. The first of these is a given/new distinction which Hawkins handles by means of the concept of sets of entities whose members are familiar to the participants in a discourse. Hawkins argues that features of the discourse context and utterances in the discourse trigger the participants to set up what he calls pragmatic sets (p-sets) of entities whose members can subsequently be referred to as familiar entities. He distinguishes between references to members of existing p-sets and references which trigger the inclusion of a new p-set. For example, in an utterance of sentence 3.20, **a cat** is said to trigger a new p-set with exactly one member, and **it** is said to refer to this p-set member.

3.20. I had a cat but it died.

The second distinction Hawkins invokes is an inclusive/exclusive distinction. An inclusive reference to a p-set member is defined as a reference to the totality of members of some p-set which satisfy the descriptive predicate of the referring expression. An exclusive reference to a p-set member is defined as a reference to some proper sub-set of the members of some p-set which satisfy the descriptive predicate. The use of **it** in 3.20 then, would be defined by Hawkins as an instance of inclusive reference. The use of **a cat** in 3.21, however, would be defined as exclusive reference.

3.21. Noah had a mating pair of every kind of animal, until a cat drowned.

In the discussion which follows we investigate the usefulness of existential features in accounting for Hawkins' range of usage types.

To return to the EF scale itself, we propose that the strongest existential effect is one which we can call **pre-establishment**. To say that an expression has the existential force of pre-establishment is to say that a hearer will understand from a speaker's use of such an expression that the existence of a referent for that expression, and the identity of that referent, were already established before the utterance of the expression[5]. In Hawkins terms, pre-establishment is a feature of p-set members.

3.2.1.1. Anaphoric Definites.

The paradigm example of an NP with the EF of pre-establishment is a non-assertive or anaphoric definite description, e.g. the highlighted definite in 3.22:

3.22. A: Who bit **the bombardier**?

B: I bit **the bombardier**.

For the interpretation of B's utterance in 3.22 as a true statement, it is necessary that there is some entity available to be identified as the referent of **the bombardier**, but the availability of this referent is not *sufficient* for the appropriate use of the definite. In addition, we propose that the existence of the entity must have been established prior to B's utterance. In 3.22, both these requirements are met, since B's utterance is preceded by A's utterance, and so B's utterance of the definite can be classified as appropriate and the statement can be analysed as potentially true. In turn, A's use of the definite, if it is appropriate, rests on the utterance of some still earlier establishing expression or the presence in the visible context of a recognisable bitten bombardier.

---

[5]The initial establishing of the existence and identity of some referent is associated with the use of expressions with a weaker level of EF and will be discussed in section 3.2.2

We must defend this analysis on three fronts. It must be shown; (1) that it is well-motivated to claim that the existential reading originates in the definite description; (2) that the EF in this sentence should be analysed using a new category "pre-establishment" rather than using the existing and widely accepted category "presupposition"; and (3) that pre-establishment merits inclusion, and inclusion at a separate level, on the EF scale.

(1) It might be suggested that the EF in 3.22 originates in the verb; i.e. the requirement for the existence of a referent for **the bombardier** results principally from its co-occurrence with the combination of the pronoun **I** and the verb **bit** rather than from any feature of the definite itself. That is, any expression which occurred here would have to be interpreted existentially. If we understand EF as a requirement of existence on the part of some entity then it does follow that it is the selectional rules for the verb **bite** which contribute the EF in this sentence. On such an analysis we would say that, for this sentence to be interpreted as an expression of a true proposition, the verb requires that its arguments have referents which physically co-exist. In addition, since the speaker of (some utterance of) this sentence necessarily exists in one particular location, this amounts to a requirement for the existence of the referent of the definite in the same location as (and furthermore in physical proximity to) the speaker.

Our answer to this objection is that it rests on a model of interpretation which is at odds with the view taken of interpretation in the current thesis. We are interested in interpretation as a dynamic process which takes place during real discourse and which relates contexts, through utterances, to further contexts. It is only from the point of view of theories such as DRT (Kamp and Reyle, 1993) or Montague Grammar (Cann, 1993), which state interpretation declaratively as a function from sentences (or their logical translations) to models, that the EF in 3.22 can be argued to originate in the verb. Put informally, the interpretation component of DRT allows the verb to dictate the features of the model with respect to which some sentence is interpreted; in this case, the model must contain a referent for **the bombardier.**

It is clear however that although a semantic analysis of the verb **bite** might select for the co-existence of the referents of its arguments, it does not follow that the *use* of **bite** in any way

confers co-existence on a pair of objects in any particular model. For example, taking sentence 3.23 as a representation of an utterance by the author of this thesis on the first of October 1995,

3.23. I danced the last waltz with **the Wizard of Oz** last night.

although **dance** here requires the physical co-presence of Catriona McPherson and the Wizard of Oz on the 30th of September 1995, hearers/readers of this sentence are not bound to interpret the sentence with respect to some model which fulfils the requirement; and if readers of 3.23 choose to interpret it with respect to the real world they cannot interpret it as a representation of a true proposition[6].

(2) The EF of definites used in the way represented in 3.23 has traditionally been dealt with under the heading of presupposition (Strawson, 1950; 1954, Karttunen, 1973, Burton-Roberts, 1989). It was argued in chapter 2, following Werth (1993), that the employment of a logical category of presupposition is misguided and unnecessary, its apparent usefulness being merely a side-effect of the tendency of linguists to examine isolated sentences out of context. We claim that when a representation of a sentence with no context is presented to a reader, as is usually the case in the literature on presupposition, it is likely that readers of the sentence immediately and automatically fill in a context to facilitate the interpretation of that sentence. We believe that the reader's lack of awareness that the process of supplying a context has been carried out means that the reader *uses* a context without acknowledging its existence or paying attention to how any particular context is chosen and so is led to mis-analyse the behaviour of definites by claiming that a definite carries an existential presupposition or triggers an existential presupposition[7]. We propose that, in fact, the relationship between definites and existence is the exact reverse of this: the use of a definite description is licensed or made

---

[6]see chapter 6 for a full discussion of choice of existential location as a part of interpretation.

[7]The unconscious supplying of contexts has been implicated in further cases of misanalysis since where there are two possible contexts for some construction the more common context tends to be assumed with the result that the construction is then defined by its occurrence in that context only. e.g. one use of **do** which typically but not exclusively occurs in sentences like **Do sit down** has largely been analysed as a marker of women's speech, since its other context, e.g. in **Do read those papers over (as I suggested earlier)** which is not predominantly used by females, is much more uncommon. (Ellen Prince, Henry Sweet Lecture , LAGB 1995)

appropriate by the existence of the object to which it refers. B's utterance in 3.22, for example, occurs after the existence of the bombardier has been offered, negotiated and accepted by the discourse participants as a result of the preceding discourse. Ideally, to avoid the pitfalls associated with evaluating sentences out of context it would be preferable to present all the example sentences in this thesis preceded by a portion of discourse to fix some features of the context. However, this is not possible due to practical constraints, but we trust that an *awareness* of the process of filling in a context will suffice.

We recognise that part of the obscurity of some of the semantic literature and much of the misunderstanding on the part of linguists working in this field of other linguists' work arises from the profusion of different terminology covering more or less the same ground. For example, presupposition is defined by different linguists as a semantic or a pragmatic relation which holds between sentences, propositions, statements, utterances, or between a speaker and one of these. The relation of presupposition has been called a weak entailment (Wilson, 1975), a semantic entailment, (Keenan, 1972), and a weak entailment without a strong entailment (Burton-Roberts, 1989). Levinson (1983) reviewed semantic and pragmatic theories of presupposition and identified a wide range of conflicting pragmatic definitions; for example, conventional implicatures (Karttunen and Peters, 1979), and the theory of potential and actual presuppositions owing to Gazdar (1979). Levinson (1983) also noted that while the category of presupposition should be reduced to other "more useful concepts", there was little consensus among linguists as to how this should be achieved (Levinson, 1983: 217). It is not, therefore, a decision to be taken lightly to introduce the new label **pre-establishment** to define what might be seen as essentially back-to-front presupposition. It would be possible to use the term presupposition with the stipulation that it is understood that hearers presuppose existence as a result of prior establishment of existence. However, the use of the term presupposition to describe a process of definites *triggering* existence is so pervasive and we find the notion of this process so misguided and misleading, that we conclude that it is preferable to use a different term.

(3) There is a sense in which if the EF scale is understood as a scale of existential commitment, and all utterances are assumed to take place in contexts, then a definite such as

the bombardier in 3.22 has existential properties which are "off the scale"; that is, the use of **the** indicates a recognition of mutual acceptance of the existence of the referent while at other points on the scale, the choice of determiner indicates the minimum necessary level of commitment or acceptance of the existence of the referent. If EF is viewed from the perspective of the speaker, as in McCawley (1981) pre-establishment would not merit a slot on the EF scale, since a speaker's level of commitment to the existence of some entity cannot be affected by whether or not the hearer shares it. However, the claim of the current thesis is that EF level is a measure of the hearer's degree of acceptance of the existence of some entity. In view of this it is valid to associate a pre-established existential proposition with a higher level of EF than some existential proposition which is undergoing consideration by the hearer for inclusion in the common ground at a particular stage in the discourse.

3.2.1.2. Pronouns.

We have argued that it is because we can analyse the uses of definites discussed above as anaphoric that we can define their EF as pre-established. Our next claim, that pronominal NPs should also be assigned a level of EF which places them at this point on the EF scale, should not therefore be difficult to defend. We claim that the pronoun **him** in 3.24, for example, must be interpreted as a reference to some object whose existence is pre-established before 3.24 is uttered.

3.24. I bit **him**.

In fact, it might be said that a pronoun is the most intuitively clear example of an NP which is appropriate only where the existence and identity of its referent are beyond debate. Indeed, if pronouns are analysed as determiners as we suggested in section 3.1.2, the existence and identity of some object to be associated with the null predicate of a pronominal NP must be *pre*-established as there is no semantic content in the predicate slot to allow an interpretation otherwise. In 3.25 (with the context outlined for 3.23) B's utterance cannot be interpreted as a

true statement since the identity of the only available referent for **him** precludes the necessary physical proximity of that entity with the speaker.

3.25. A: What can you tell me about the Wizard of Oz_i?
B: I danced the last waltz with **him**$_i$ last night.

It is not our intention to suggest that anaphoric pronouns which are coindexed with discourse internal, linguistic antecedents are in any way more basic than deictic pronouns whose referents are fixed directly by features of the context.

3.2.1.2.1. Basic and Derived Pronoun Types.

Some procedural semantic frameworks do take anaphoric pronouns with linguistic antecedents as prototypical pronominals and handle deictic pronouns by analogy. For example, in Heim's file-change model (Heim, 1982), referring expressions are associated with file-cards. She proposes that an indefinite triggers the introduction into the universe of discourse of a new file-card, while a definite brings about the modification of some existing file-card by the addition to it of more information concerning the entity with which it is associated. Anaphoric pronouns, in Heim's analysis, are always linked to existing file-cards. Deictic pronouns, on the other hand, are handled by the inclusion in the file of an initially blank context file-card onto which pseudo-antecedents for deictic pronouns are introduced by an accommodation mechanism which is activated at the point in the discourse when the deictic pronoun is uttered. Deictic pronouns are then co-indexed with these "antecedents" in a way which mirrors the association of anaphoric pronouns with linguistically expressed antecedents.

We do not agree with Heim that it is a well-motivated step to analyse anaphoric pronouns as more basic than deictic pronouns. In fact, there may be some evidence from child language acquisition and from historical and cross-linguistic studies, for the opposite claim: that deixis is more basic than anaphora.

The evidence from child language acquisition on this point is indirect. Children acquire deictic function relatively early in their language development far in advance of their acquisition of anaphoric function (Bowerman, 1973; Radford, 1990). For example, Crystal (1987) quotes Stoel-Gammon and Cooper (1984) who list "*What's that?*" among the first fifty utterances of two children whose lexical development was studied between the ages of eleven and sixteen months (Crystal, 1987).

However, it is also a feature of child language acquisition that children begin with what Bloom, Lightbown and Hood call a nominal style which persists until the relatively late development of the determiner system, including personal pronouns (Bloom *et al.*, 1978). For example:

3.26.	Amy draw boat.	(I am drawing a boat)
3.27.	Mummy tickle Amy	(Tickle me)
3.28.	Amy touch Amy toes	(I am touching my toes)

Abney (1987) argued that the lack of any overt inflections on *this* and *that* in English allows children to misanalyse them as full nominals rather than pronominal determiners. So it appears that although utterances of *that* should be analysed as having deictic function they cannot be analysed as having the form of pronominal determiners.

Clearer evidence that deixis is more basic than anaphora comes form the argument that anaphoric function actually derives from deictic function (Lyons, 1975; 1977).

> "... deictic distinctions can be used to identify the antecedents of anaphoric expressions. Anaphora involves the transference of what are basically spatial notions to the temporal dimension of the context of utterance and the reinterpretation of deictic location in terms of what may be called location in the universe-of-discourse."
>
> (Lyons, 1977: 670)

The view that anaphoric function has derived from deictic function in the evolution of language across time is well-motivated in that it conforms to the general trend of semantic

change identified by Traugott (1982, 1989). Traugott proposes three distinct semantic components in language: the propositional component, consisting of the means by which a language connects with states-of-affairs, e.g. concrete nouns: *table* and action verbs: *kick*; the textual component, consisting of the means by which discourses are organised, e.g. textual connectives: *however* and anaphoric devices: *the former*; and the expressive component, consisting of the means available to a speaker for expressing attitude, e.g. sentential adverbs: *surprisingly*, *unfortunately*. Traugott claims that semantic change tends to involve process whereby propositional meanings develop into textual meanings and then into expressive meanings. McMahon (1994) gives the example of Old English *þa hwile þe* 'at the time that', which links two propositions, developing first into Middle English *whyle* 'while', which can also be used to link two clauses, and then into present-day English *while* which, besides having these functions, can be used with the meaning of 'although' to link two attitudes on the part of the speaker (McMahon, 1994: 187). For example:

3.29.	While I was sleeping, someone stole my car.	Propositional use
3.30.	While the lioness is fierce, the lion is placid.	Textual use
3.31.	While I enjoy the arias, some of the plots are laughable.	Expressive use

If Traugott's claims about general trends in semantic shifts are correct, then the development of anaphoric, textual functions associated with pronouns whose original function was to express propositional, deictic meanings is unremarkable.

If we accept that the anaphoric use of a pronoun is historically derived from its deictic use, can be analysed as a special type of deictic use, and is acquired later than its deictic use, then any analysis which treats anaphoric function as more basic than deictic function is called into question at least at a pre-theoretical level, and an analysis which derives anaphoric uses from deictic uses would be preferred. However, from the point of view of the interpretation process the most satisfactory analysis would be one in which neither aspect of the total context of utterance, whether linguistic or extra-linguistic, was seen as basic. The reason for this is that, as was shown in Werth (1993), it does not appear that the method by which the proposition that some entity exists enters the common ground affects the subsequent potential for

pronominal references to it. For example, referents for **it, he**, and **she** are necessary for the proper use of these pronouns in sentences 3.32 - 3.34, but there is no difficulty in interpreting the sentences even though the existence of suitable entities must be inferred from the preceding linguistic context.

3.32. I'm pregnant so **it** will be born in the spring.

3.33. I was married before, but **he** was a swine.

3.34. I wanted to get my hair cut but **she** was fully booked.

It is possible to cancel the association between **it** and the baby, **he** and the husband, and **she** and the hairdresser as sentence 3.35. shows:

3.35. I was married before but **he**$_i$ was a swine, although my husband$_j$ was alright.

The important point is that it is not possible to cancel the association on the grounds that the entity does not exist. If the most likely referent is explicitly ruled out then some other referent must be available, e.g. sentence 3.36.

3.36. I was married before but **he**$_j$ was a swine - my father-in-law$_j$ that is, not my husband$_i$ - so this time I'm going for an orphan.

It should be noted, however, that the use of **he** in 3.36 is not possible without previous explicit mention or some other type of contextual availability of the speaker's father-in-law.

3.2.1.3. Names.

The third type of NP we propose for inclusion at this level of EF is the class of proper names. We propose that the object in 3.37 exhibits the same EF as the definite in 3.22 and the pronoun in 3.25.

3.37. I bit **Bilbo**.

Our claim is that 3.37 is appropriately used only in those contexts of utterance where there is an entity which is identifiable as the referent of the name **Bilbo,** and the existence of the entity and it's being named Bilbo are pre-established.

One possible objection can be identified to our inclusion of names at the same level of EF as definites and pronouns. The argument in this objection rests on a view of the semantics of names in which names are analysed as rigid designators which *must* refer and always refer to the same entity. On this analysis, names only cause sentences to be uninterpretable, if, as is the case if we try to interpret 3.38 (taking the real world as the model), the proper names do not have any reference.

3.38. **Coynburgh** is the county town of **Tarelginshire**.

We must supply justification for the added requirement that the existence of a referent for a name be pre-established. Take for example sentence 3.39, in the current context (i.e. in which 3.39 is a token of a sentence written by the current author and read by the current reader):

3.39. **Brian** is three years older than **Andrew**.

It may be the case that this sentence is as uninterpretable as sentence 3.38 or that the referents for the names can be identified and the truth-value of the sentence assigned. Which of these two possible situations actually holds depends on the current reader's access to the appropriate context.

In expositions of Montague Grammar or DRT, where the context to be used in interpretation is fully defined by the model theory, any example sentence can be interpreted, since it is always clear whether the specified index contains the required entities in the required relations. In other words, in Montague Grammar or DRT, interpretation is an idealised concept, abstracted away from the utterance context, generalised to cover both speakers and hearers, and not constrained by any epistemic limitations on the hearer's access to the correct index in the model. However, it is a central contention of the current theory that a crucial part of the interpretation of utterances consists in locating the referents of all referring expressions at the correct index in a model which, for any individual hearer, is incompletely specified.

We can make the stronger claim about the appropriateness of names, because we accept that for some hearer of 3.39 who cannot identify the named entities, 3.39 is as uninterpretable as 3.38 which contains names without any referents. In other words, we are claiming that it is not of primary importance whether or not referents for names *exist*, but that if hearers do not know that they exist, then they will be unable to interpret sentences containing the names[8].

We also propose the inclusion of the determiner GEN at this level on the EF scale. However, we cannot offer an argument in support of this decision until after the discussion of the next lowest EF level. We will now turn our attention to this discussion and return to the question of the EF of generics in section 3.2.2.5.

3.2.2. Assertion

3.2.2.1. The EF of Indefinites.

The second highest level on the EF scale is one which we can define as being associated with assertions of existence by the speaker. At this level we can place those determiners which are used to assert the existence of entities for the first time in a discourse. The idea of assertive

[8]The possibility of weaker EF for some uses of names is discussed in section 3.2.2.4.

EF seems to accord rather well with people's intuitions about many uses of the indefinite articles **a/sm** which Hawkins (1978) would categorise as references to entities which are not members of any existing p-set but trigger the introduction of a new p-set. For example;

3.40. **A student** tried the meatloaf recipe.

3.41. **Sm monkeys** have landed the contract to run the zoo.

Traditional analyses have viewed the EF of **a/sm** as *entailment* and contrasted it with *presupposition* arguing that these two contrastive options constitute the entirety of positive existential features of meaning (Kempson, 1977). We have already made it clear that we believe it is useful to recognise a richer range of existential levels of meaning. In addition, the formulation of our theory of the interpretation of referring expressions is at odds with the characterisation of the EF of an indefinite as a kind of entailment. To see the EF of **a** as an entailment, it is necessary to take a view of interpretation which works from sentences to contexts, that is, to employ a model theoretic interpretation procedure. Using such a procedure it is plausible to argue that the appropriate utterance of sentence 3.42, for example, entails that the model in which the interpretation of 3.42 is embedded, must contain at least one individual who is a retired lion tamer and who lives in or comes from Cardiff i.e. the use of the indefinite entails the existence of a referent *in the model*.

3.42. **A retired lion tamer from Cardiff** has won first prize in this week's national lottery.

However, our approach views interpretation in a less idealised way from the point of view of the hearer. This has the effect of turning the relation around. Instead of an entailment from the indefinite to the lion tamer, we express the connection between the indefinite and the lion tamer as a type of licensing relation. That is, there must be at least one retired lion tamer who lives in or comes from Cardiff for the speaker of some utterance 3.42 to use it to make a true statement. As we pointed out in section 3.2.1.1, our analysis has the result that, from the point of view of the speaker there is no distinction between the requirements for the utterance of expressions with pre-established EF and asserted EF; in both cases, the existence of some

object must be known to the speaker before an utterance of some sentence containing either a definite or indefinite reference to that object can be used to make a true statement[9].

From the point of view of the hearer, however, while prior knowledge of the existence of some object is necessary to interpret properly an NP with pre-established EF, NPs with what we are calling asserted EF do appear to entail the introduction and therefore the existence of their referents in the knowledge base of the hearer. This rests on the operation during any discourse of a mutually acknowledged co-operative principle (Grice, 1975) on the strength of which hearers will accept statements made by speakers and update the common ground of the discourse according to reported information. We acknowledge that the negotiation of an existential proposition into the shared context can be initiated by the utterance of an indefinite. If the hearer of 3.42, for example, chooses to accept the addition of the individual referred to by the indefinite to the set of individuals which exist in the shared context, then we can say that the establishing of a lion tamer in Cardiff in the knowledge base of the hearer is entirely due to the occurrence of the indefinite description.

What this amounts to is that when hearers assign truth-values to utterances, and specifically when they assign the truth-value *true*, they very often do so on solely on the basis that, according to the co-operative principle, if the speaker has asserted that some state of affairs is the case then it must be the case. However, we believe that although indefinites *may under certain conditions* trigger the acceptance of the existence of an entity on the part of a hearer, and that this could be seen as a kind of pragmatic entailment, this does not offer any motivation to analyse assertive EF in terms of semantic entailment. We must evaluate the proposal in Hawkins (1978) that assertive EF can be analysed as pragmatic entailment.

Hawkins' claim that this use of indefinites triggers the introduction of a new p-set can be seen as equivalent to a claim that this use of an indefinite pragmatically entails the existence of its referent (Hawkins 1978, 1984). We propose that defining assertive EF in terms of even *pragmatic* entailment is misguided and misleading.

---

[9]the existential status of the referents of definites and indefinites in opaque contexts will be discussed in chapter 7.

The difficulty with arguing against a pragmatic entailment analysis is precisely the fact that the co-operative principle ensures that most example sentences containing indefinites will be interpreted as counter-examples. However, there are cases in which the appearance of a pragmatic entailment relation can be cancelled. These are cases where the hearer's existing epistemic state rules out the acceptance into the shared context of some entity which is offered for negotiation into the context by an indefinite referring expression. For example,

3.43. I saw **a ghost** last night.

3.44. There are **sm fairies** at the bottom of my garden.

We claim that these examples show that rather than expressions with assertive EF pragmatically entailing the existence of their referents, the existence of referents is a pre-requisite for the use of these expressions in utterances used to make true statements. We assume that readers of 3.43 judge it to be false in the current context and that its perceived falsity is caused by the inability of readers to accept the existence of a ghost which the writer of 3.43 has seen. If the use of an indefinite really did pragmatically entail the existence of its referent for a hearer there would be no easy way to explain the perceived falsity of 3.43.. *However, if only the availability of a referent can license the use of an indefinite, it follows* from the unavailability of a referent for **a ghost** in 3.43, that the use of an expression with assertive EF is not appropriate, that the indefinite description is not licensed and that, therefore, 3.43 cannot be used to make a true statement.

As discussed at the start of this chapter, we propose that bare plurals be analysed using phonologically null determiners and since the two sentences in 3.45, for example, appear to be exact paraphrases it is proposed that some occurrences of bare plurals are assigned the null determiner **SM**, equivalent to the phonologically realised determiner **sm** which has assertive EF.

3.45a. **Sm monkeys** have seized control of the zoo.

3.45b. **Monkeys** have seized control of the zoo.

3.2.2.1.1. Other Uses of Indefinites.

We stated at the beginning of this section that assertive EF characterises *some* of the uses of indefinites; those which Hawkins categorises as instances of reference to entities not included in any p-set. The remaining uses of indefinites are those which Hawkins classifies as instances of exclusive reference to members of existing p-sets, e.g. the use of **a cat** in 3.21 (repeated here).

3.21. Noah had a mating pair of every kind of animal until a cat died.

Recall that we noted in chapter 2 section 2.1.2 that one important advantage of Russell's analysis of indefinites as existential conjunctions was that this allowed us to rule out the possibility of a sentence such as 3.46 being assigned the truth-value *true* if its speaker could continue with the second conjunction of 3.47.

3.46. A black cat didn't cross my path.

3.47. A black cat didn't cross my path but another one did.

However, it is not accurate to rule out *all* sentences of the type shown in 3.47. There are numerous examples, such as 3.48, which are perfectly acceptable and can be interpreted as potentially true statements.

3.48. The doctor thought I might have thrombosis since I had a swollen leg, but she changed her diagnosis when she realised that the other one was all right.

We can show that by invoking Hawkins distinction between the two types of indefinite use, we can predict why the conjunction in 3.48 is acceptable but the conjunction in 3.47 is not. Werth (1993) claimed that the means by which propositions enter the context of a discourse are irrelevant to their subsequent status as part of the common ground. Werth went on to argue that any existential proposition in the common ground could be treated as backgrounded

information i.e. could be associated with a definite. We argued, following on from Werth's observations, that only those propositions which are either part of the common ground or are being offered for inclusion in the common ground can be negated.

If we analyse the use of the indefinite in 3.49 as assertive (what Hawkins calls reference to a entity which is not in any existing p-set), then we can explain why the existence of this entity cannot be in the scope of the negation in sentence 3.46 since this would involve the speaker in a contradiction; both asserting and denying the existence of an entity.

3.49. A black cat crossed my path.

Furthermore, if the indefinite in 3.46 is assertive, then the continuation in 3.47 is not possible because there is no referent available for the pronominal expression **another one**. The assertion of the existence of *one* black cat in the context, does not facilitate the location of any *other* black cat in the context. In contrast, in sentence 3.48, a set of legs with two members associated with the speaker is part of the common ground of the discourse as a result of general background knowledge about human physiology. Therefore, a **swollen leg**, in this context, *can* be interpreted as an instance of exclusive reference to a proper sub-set of a p-set of legs. This explains why a referent is available for the expression **the other one** in this sentence. Since the existence of the set of legs in this context is part of the common ground, the speaker would be able to deny the proposition that either or both of these legs exist(s) without uttering a contradiction. For example:

3.50. The doctor thinks I may have thrombosis, since I have a swollen leg.
The other one is not swollen - it was amputated years ago.

We could account for the possibility of denying the existence of an entity picked out by an indefinite description if we analyses the indefinites Hawkins associated with exclusive reference as having pre-established EF. There is some precedent for positing a mismatch between the pair of categories we are calling pre-established and assertive EF, and the pair of expression types exemplified by **the** and **a/sm** (Hawkins, 1978; Sperber and Wilson, 1986;

Werth, 1993) but usually the claim that is made is that definites have some assertive uses, not that indefinites have anaphoric uses. In the next section, we will consider in detail the so-called assertive uses of definites, pronouns and names to see if they coincide with the uses of indefinites which we have claimed might be more properly analysed as non-assertive or anaphoric. If this is the case, then the motivation for analysing such definites, pronouns and names as assertive is greatly reduced.

3.2.2.2. Definites.

We indicated in the discussion of anaphoric definites that it might be possible to analyse some definites in another way. There are three interesting treatments of definites in the semantic literature, reviewed below, which all make similar distinctions between anaphoric and what we might call assertive definites. These theories make the common claim that assertive uses of definites are restricted in a way that the uses of other assertive expressions, such as indefinites are not. We will evaluate this claim and then consider an alternative analysis which argues that, because these restrictions are identical to the restrictions on uses of indefinites which are better analysed as non-assertive, it is not possible or well-motivated to claim that any definites are assertive.

Werth (1993) distinguishes anaphoric definites, which he claims contain only given, background existential propositions, from non-anaphoric definites, which are used to make unconventional assertions of new existential propositions[10]. He argues that the degree of appropriateness of unconventional assertions correlates with the level of expectedness of the asserted existential proposition.

On this point, his analysis shows similarities to the distinction Sperber and Wilson make between foreground and background entailments (Sperber and Wilson 1986), in which it is argued that foregrounded existential entailments, which are expressed by indefinites, may

[10]See the review in chapter 2, section 2.5.

contain information which is costly in terms of processing effort, whereas backgrounded existential entailments, which are reflected in definites, must involve little processing cost.

Both of these approaches are heavily indebted to Hawkins' account (Hawkins 1978) of the semantics of definites and indefinites in which he argues that assertive definites are only appropriate if they refer to some entity whose existence in the common ground i.e. in a p-set predictable on the grounds of stereotypical expectations.

Table 3.3 shows the similarity among the three approaches.

Table 3.3
Restrictions on anaphoric definites

	anaphoric definites	**assertive definites**
Hawkins (1978)	second mention definites licensed by antecedents in the shared previous discourse set	first mention definites licensed by membership of some other shared set
Sperber & Wilson (1986)	anaphoric definites reflecting backgrounded entailments	assertive definites expressing fore-grounded entailments
Werth (1993)	anaphoric definites referring to given information in the common ground	assertive definites conveying uncontroversial new information

Hawkins first divides the class of definite usages into two broad categories: first and second mentions. Second mentions can be classified as anaphoric. The highlighted NP in sentence 3.51, for example, shows anaphoric usage in which a definite is coindexed with a preceding indefinite in the utterance:

3.51. A man$_i$ came in, and then a woman and then, I think, a another woman might have. I definitely remember **the man$_i$**.

Hawkins also includes a range of cases in which no straightforward co-indexing is possible but which can nevertheless be analysed as instances of indirect anaphora. The highlighted definites in 3.52 and 3.53 are examples of some of these types of usages.

3.52. I saw a man and a woman sitting on a bench. **The couple** were obviously having an argument.

3.53. I don't want to drive to London again. **The journey** is easier by train.

The last and most interesting class of anaphoric definite usages is what Hawkins calls associative anaphora. These are instances of definites which refer to objects whose existence is in some way implied the preceding discourse. The highlighted definite in 3.54 is an example of an associative anaphoric use in which the definite is licensed because the presence in the common ground of a church implies the presence in the common ground of a set of entities which co-occur with churches; e.g. ministers.

3.54. I went to a church and couldn't find **the minister**.

Associative anaphoric uses are particularly interesting because their appropriateness seems to depend on hearers' beliefs about reliable stereotypes. Hawkins handles this feature of their meaning by employing a stereotype requirement, which marks as deviant sentences such as 3.55, in comparison with 3.54, since the assumption of the presence in the common ground of a giraffe, after the mention of a church, is not licensed by our background knowledge of stereotypical churches.

3.55. I went to a church and couldn't find **the giraffe**.

The example in 3.56 was uttered by a six-year-old while taking part in Christmas celebrations at primary school *for the first time*:

3.56. I'm in the nativity play - I'm one of the Hawaiian dancers.

The speaker in this example inadvertently flouts the stereotype requirement, since with only one example of nativity plays to draw on, she cannot distinguish stereotypical components from others.

As well as these types of anaphoric uses, Hawkins distinguishes among several varieties of first-mention usages, which cannot be interpreted in any way as linguistically anaphoric, but depend for their interpretation on either the visible situation, the immediate non-visible situation or the larger situation. For example, sentence 3.57 uttered by a speaker to a hearer while both are watching a ballet, contains a first-mention definite where the visibility in the physical situation of utterance of a suitable referent for the definite fulfils the requirement for availability and identifiability.

3.57. **The ballerina** looks tired.

Hawkins uses the highlighted definite in 3.58 (on a sign on a garden gate) as an example of a first-mention usage which is licensed by the immediate situation.

3.58. Beware of **the dog**.

Like associative anaphoric uses, this usage is said to be subject to a stereotype effect. In this example it is argued that it is our knowledge that houses regularly have dogs in them which licenses the use of the definite. However, this particular example does not perhaps provide sufficiently strong evidence for the stereotype effect, since "beware of the dog" is a fixed idiom in the sense that replacing **the dog** with any other definite results in a bizarre seeming sentence whether or not the substituted expression refers to an entity of a type regularly found in houses or not. For example, 3.59 and 3.60 seem equally odd; neither is completely anomalous but neither is as acceptable as 3.58.

3.59. Beware of **the cat**.

3.60. Beware of **the giraffe**.

However, it is possible to identify a range of examples which do support the claim that a stereotype effect is at work in first-mention uses, although it is not entirely clear whether these are examples which Hawkins would analyse as immediate situation uses or larger situation uses (see below). Consider sentences the two sets of possible responses to the question in 3.61.

3.61. Did you have a good Christmas?

a. Not really. The pipes froze. The boiler burst. The chip pan went on fire on Boxing day. The dog ate the turkey and the car got stolen.

b. Not really. The giraffe died and the photocopier never turned up.

In the response a. all the definites are easily interpretable as immediate situation uses, licensed by our knowledge that people often spend Christmas at home in houses which typically contain pipes and boilers, and regularly contain dogs, cars and chip pans[11]. In contrast, giraffes and photocopiers are not *stereotypical* features of houses. The b. response, therefore, appears peculiar when presented out of context in a way that the a. response does not.

The definites which Hawkins identifies as larger situation uses are those which are dependant on the uniqueness of their referent relative to some situation which is not defined by the discourse but is larger than the physical situation and possibly smaller that the world as a whole. For example, the highlighted definites in 3.62 and 3.63 could be analysed as examples of larger situation definites.

3.62 **The Prime Minister** has announced her resignation.

3.63 I'm not going to **the nine-o'clock lecture** tomorrow.

---

[11]Note that **the turkey** would be analysed as an associative anaphor of **Christmas.**

A token of 3.62 was broadcast on BBC R4 in November 1990 and in this context, the UK, since there is an entity which uniquely satisfies the descriptive predicate of the definite, this use is licensed by the availability and identifiability of a referent. This use would not have been appropriate on a BBC World Service broadcast, for example, since the identity of the intended referent would not be uniquely fixed by the definite in that context, where more than one female Prime Minister exists. Likewise, the highlighted definite in 3.63 is appropriate between university class-mates, since there can only be one nine o'clock lecture each day in each course.

Although we agree with Hawkins' observations and acknowledge that definites can be handled by distinguishing the two types of usage he proposes, we cannot conclude that a split between anaphoric and assertive definites is necessarily the only, or the best, way to account for the data. Rather than accepting that there is a category of assertive definites which is subject to stereotype restrictions, we could argue that there is no need for this category if by taking anaphora to include a very broad range of relations, we classify all definites as anaphoric.

We will examine several aspects of the question of whether to split definites into assertive and anaphoric uses. First, we will try to determine the position of the putative division. Secondly, we will consider whether by accepting the existence of such a split between assertive and anaphoric definites we weaken other areas of our emerging characterisation of the EF scale. Thirdly, we will examine the evidence for the two competing analyses.

3.2.2.2.1 The Position of the Dividing Line Between Assertive and Anaphoric Definites.

Hawkins identifies three categories of definite usage which, he argues, employ anaphoric means: direct, indirect and associative anaphoric uses. If we could support an argument that the remaining three uses, visible, immediate and larger situation uses, are in any way more assertive than each of the first three categories we could confidently claim that what distinguishes anaphoric from assertive definites is whether the conditions which license their

use are linguistic or non-linguistic. However, if we consider examples of each type of definite in terms of how assertive *seeming* they are, we find that the intuitive distinction does not accord with the linguistic/non-linguistic distinction.

Table 3.4

Conflicting definitions of EF for definites

	Linguistically licensed	**Contextually licensed**
Anaphoric ?	I saw a man. I bit the man. I swore at a policewoman. The oath appeared to offend her.	The chorus sounds tired. (at the opera)
Assertive ?	I went to the church. I couldn't find the minister.	Beware of the dog. The Prime Minister ate his peas.

The examples of direct and indirect anaphoric use seem unequivocally non-assertive, as does the visible situation use and this instance of a larger situation use, although other examples such that in 3.63 appear less so. In contrast, the associative anaphoric use and the immediate situation use appear more assertive. However, even though there does not appear to be strong evidence of a clear distinction, we could argue that there are clear cases which accord with the linguistic/non-linguistic distinction and that we could pay the price of treating the others as exceptional if there were theoretical as well as this rather ambivalent observational motivation for positing the split.

3.2.2.2.2. The Effect of an Assertive/Anaphoric Distinction.

It is far from being an unprecedented step to propose a distinction between the contributions to the interpretation process of the linguistic and extra-linguistic context. It might even seem unaccountably squeamish to hesitate at all over employing this tactic. It is a general feature of many semantic theories that there is a dichotomy between the preceding discourse and the more general context of utterance. It is uncontroversial to categorise anaphoric and deictic pronouns separately and Hawkins' analysis of definites into first and second mentions relies on the acceptance of this kind of distinction. Formal theories very often take the further step of ignoring entirely the non-linguistic context. Such a decision can be justified for sentence-level model theoretic frameworks such as Montague Grammar but this type of division is also employed in current formal *discourse*-level approaches to first-mention definites as the following quotation from Kamp and Reyle's exposition of DRT (Kamp and Reyle, 1993) shows.

> "The initial context k_0 for a given discourse should be a DRS incorporating that (relevant) information which is available to the recipient of the discourse at the point when he starts processing it. *Nothing of importance is lost, however, if we make the simplifying assumption that the initial context contains no information at all.*"
>
> (Kamp and Reyle, 1993 (emphasis CMcP))

The result of this assumption is that none of Hawkins' class of first mention definites can be accounted for in Kamp and Reyle's framework. Clearly, we could not take the step of ignoring the context all together, but we must ask ourselves what would be the result for our emerging analysis if we were to make the weaker assumption that we can distinguish anaphoric and assertive definites along the linguistic/contextual split and account for them in different ways. This would involve anaphoric definites being analysed as referring to objects whose existence is pre-established and assertive definites being analysed as referring to objects whose existence is expected or uncontroversial. Recall that we claimed that by accepting Werth's (1993) argument that some definites should be analysed as assertions, we could avoid the problem of ambiguous negation, since only in utterances in which the existential

propositions were not being asserted would it be possible for them to be negated. It would only be possible to continue to hold this assumption, and to distinguish assertive and anaphoric definites, if it could be shown that so-called assertive uses cannot be negated and that this results from the impossibility of simultaneously asserting and negating the same proposition..

A simple example can establish that for some utterances of negative sentences containing assertive definites which are contextually licensed no interpretation is possible if it is the existential proposition which is the focus of negation. For example, if 3.64 is uttered in a situation where exactly one cat is visible, and **the cat** is interpreted as a reference to that cat, the sentence cannot be assigned the truth-value *true* on the basis of the truth of 34a..

3.64. The cat does not belong to me.

3.64a. It does not exist.

However, even though 3.64 on this reading is not true we cannot accept that this is a result of the fact that the speaker has contradicted her own assertion. It seems obvious that 3.64 will be judged to be false because the cat's existence is physically established and that the use of the definite merely reflects this. We conclude for visible situation uses, the existential propositions associated with definites cannot be negated but that this is not the result of the impossibility of simultaneous assertion and negation of the same proposition. This example does not allow us to conclude that definites are assertive.

If we turn our attention to immediate situation uses, we find examples which suggest that the existential propositions associated with definites *can* be negated. Consider sentence 3.65 in a context where the speaker is calling to a stranger who has been knocking at their next-door neighbour's front door and is now walking round the side of the house.

3.65. The back door isn't open.

3.65a. It's padlocked.

3.65b. There isn't one.

It *is* possible to negate sentence 3.65 (on the b. reading), but we could argue that the proposition that the back door exists is implied by the actions of the hearer not by the speaker's use of the definite i.e. it is possible to argue that the use of **the back door** in 3.65 is not assertive. Stronger evidence still can be found in immediate situation uses where the existence of the referent of the definite is necessary rather than merely stereotypical since we find in these cases, such as 3.66, that negation is not possible.

3.66. Alcoholism has not ruined the liver of this patient.

3.66a. In fact, his liver is remarkably healthy.

3.66b. This patient does not have a liver.

Although it is not possible for 3.66 on the b. reading to be a true statement, we must attribute this to the impossibility of a human being without a liver. This example cannot be used as proof that there are such things as assertive definites which resist negation.

An examination of larger situation uses reveals similar findings. Sentence 3.67 on the b. reading is not possible but this is because we know that the Prime Minister exists and that therefore a negation of the proposition that he exists is false.

3.67. The Prime Minister has not died.

3.67a. But he is critically ill.

3.67b. There is no such person.

We conclude that if the defining characteristic of an asserted proposition is that it cannot be negated, then there is no evidence from the three first-mention uses of definites we have considered that they are assertive because in each case either they can be negated or the impossibility of negation must be analysed as originating elsewhere.

3.2.2.2.3. Evidence For and Against the Assertive/Anaphoric Distinction.

We have claimed that if the stereotype requirement exists this allows two possible analyses of definites and indefinites. We will examine the case for each of the two analyses in turn.

The strongest case for an analysis which distinguishes anaphoric from assertive definites would be if we could show that definites are not, in fact, restricted by a stereotype requirement; that is, if definites can be used to assert not only *uncontroversial* existential propositions but also controversial propositions, as these uses could not be analysed as anaphoric in any way.

Ideally, we could reject this analysis by showing that definites which contravene the stereotype requirement are uninterpretable or lead to a breakdown in discourse. In our framework, discourse breakdown would be signalled by the hearer's inability to assign a truth-value. However, we can predict that whether or not the analysis is correct, we will not be able to show that the result is that the discourse breaks down whenever the stereotype requirement is flouted, since discourse participants, behaving co-operatively, are normally able and willing to carry out quite substantial amounts of on-line repair and remedial work on discourses (Brown and Yule, 1983). To formulate an argument in support of the stereotype requirement we must, therefore, find a way of by-passing the confounding influence of the co-operative principle[12]. We propose the following test.

We assume that if a speaker erroneously assumes that the hearer has some piece of knowledge and makes an assertion which rests for its interpretation on the hearer's using this piece of knowledge then the hearer is entitled to draw attention to the speaker's error and request the missing information. For example, consider the small discourse in 3.68.

3.68. A: I deny categorically that I ever did so and therefore ...
B: Wait a minute! Did what?

---

[12]see the discussion in chapter 4 section 4.5

We will apply this "What? test" to the definite uses we have considered so far.

Anaphoric:

3.69. A: A man$_i$ came in, and then a woman and then, I think, a another woman might have. I definitely remember **the man$_i$**.

B: What man?

Indirect anaphoric:

3.70. A: I was behind a woman and her husband at the supermarket. The couple were having a raging argument.

B: What couple?

Associative anaphoric:

3.71. A: I went to a wedding where the bride wore black.

B: What bride?

Visible situation (context as for 3.57.):

3.72 A: The ballerina looks tired.

B: What ballerina?

Immediate situation (context as for 3.65):

3.73. A: The back door's locked.

B: What back door?

Larger situation:

3.74. A: The Prime Minister didn't eat his peas.
B: What Prime Minister?

These sentences all fail the What? test in that it does not appear that B's response in any of these cases can be interpreted as a legitimate request for missing information. In contrast, in 3.75 and 3.76, although the definites are not uninterpretable, i.e. although the hearer can add the existential propositions required for their interpretation, we claim that the addition of these propositions should count as on-line repair of the discourse, a view which is supported by the fact that these definites pass the What? test i.e. the repair of the discourse can be undertaken overtly.

3.75. A: I went to the church but couldn't find the giraffe.
B: What giraffe?

3.76. A: The Pharaoh has died.
B: What Pharaoh?

We can compare these with instances of uncontroversially assertive expressions which are not constrained by the stereotype requirement. What? questions in such cases can only be interpreted as requests for *further* information, not *missing* information, e.g. 3.77 and 3.78.

3.77. A: I have adopted a marmoset.
B: What marmoset?

3.78. A: A microwave sandwich toaster fell down my chimney last night.
B: What microwave sandwich toaster?

If we accept that such uses as those in 3.75 and 3.76, which would surely flout any stereotype requirement, can result in a potential breakdown of the discourse, we can argue that although

the discourse may recover, we are not bound by our intention to account for definite usage to be able to analyse them. Therefore, we can conclude that all the definite uses we must account for can be analysed using the notion of a stereotype requirement.

A strong case can be made for the second possible analysis of definites, on which all definites are treated as anaphoric, by considering the occurrences of pronouns in the set of environments proposed by Hawkins for definites. We will show that for each of the six definite usage types there is a corresponding pronoun usage type. Using these observations, we could either propose that there is a category of assertive pronouns or, assuming that pronouns cannot be analysed as assertive (a claim which can be backed up by applying the What? test), that the occurrence of definites in any of the six environments does not force us to analyse them as assertive either.

3.2.2.3. Linguistically Licensed Uses of Pronouns.

No argument is required to support the claim that pronouns can occur in straightforward co-referential positions. We show in examples 3.78 and 3.79 below, that the environments Hawkins analyses as indirect anaphoric uses are also available to pronouns.

3.78. Mary got married last year. He's a dentist.

3.79. I saw a man and a woman sitting on a bench.
They were having a raging argument.

There is a discussion in Kamp and Reyle (1993) of the type of anaphora in 3.79 which they handle by means of what they call a summation operation, by which the discourse referent of **a man** is added to the discourse referent of **a woman** to obtain a third discourse referent which can serve as a straightforward antecedent, thus allowing **they** to be interpreted as a regular anaphoric pronoun. Summation can also be applied to antecedents, like that in 3.80, which have to be inferred from the linguistic context:

3.80. My neighbour and I are both pregnant so **they** will be in the same class at school.

Sentences 3.81. and 3.82. represent possible examples of pronominal associative anaphora.

3.81. I went to the beauty salon but **she** was fully booked.

3.82. When I got locked out of my flat I tried to get help at a fire station but **they** were all out on an emergency call.

It appears that these sentences are as acceptable as the examples of associative anaphoric definites such as 3.54 (repeated here).

3.54. I went to church but couldn't find the minister.

We conclude therefore, that pronouns have the same range of linguistically licensed uses as definites.

3.2.2.3.1. Contextually Licensed Uses of Pronouns.

As well as the accepted deictic uses of pronouns, which correspond to Hawkins' visible situation uses of definites, we can identify immediate and larger situation occurrences. Sentence 3.83. would be an acceptable utterance from a neighbour to a stranger knocking on the door of the next house.

3.83. He's out.

Likewise, sentence 3.84. could be appropriately uttered by a speaker to a stranger in the street who was looking in exasperation at a car which was boxing in her own.

3.85. She's just gone into the dry-cleaners.

The intuition is that these pronouns do not appear assertive on the What? test, i.e. what? questions cannot be analysed as requests for missing information.

Larger situation pronouns are not common, since there are few contexts in which a unique referent can be identified on the grounds of gender alone; however sentences 3.86 and 3.87 would be plausible national newspaper headlines in the given contexts.

3.86. It's over. (the royal marriage)

3.87. He's innocent. (OJ Simpson)

In summary, we cannot claim that every definite usage type is as unequivocally anaphoric seeming as the cases where the definite is directly co-indexed with a co-referential indefinite antecedent. However, there does not appear to be sufficient evidence or other motivation for splitting the set of definites into two types; assertive and anaphoric. The contexts in which we find the "more assertive" definites coincide with the contexts in which we find the "least assertive" indefinites. If we chose to ignore this and analyse first-mention definites as assertive, we would have to acknowledge that the existential assertions made using definites are constrained in a way that other existential assertions made using indefinites are not, and we would have to account for this. Furthermore, since the environments where definites occur are also available to pronouns, we would not be able to attribute the definites' assertive status to their occurrence in these environments, but would have to find some other explanation. For these reasons we conclude that it is preferable for all definites to be analysed as having pre-established EF, to take all definites to be anaphoric on a broad understanding of what constitutes anaphora.

3.2.2.4. The Uses of Names.

Having decided that definites are to be associated exclusively with pre-established EF and restricted to the usage types discussed, we will now consider the question of the assertive potential of names. We will first test the six definite environments to see if names can occur in each of them. We will then try to establish whether the use of names is restricted to these environments or whether names can be used to assert the existence of entities more generally.

Consider sentence 3.88 in which a name occurs in what, for definites and pronouns would be a straightforward anaphoric context.

3.88. I met a friend on the train. Graham was on his way to an interview.

It is possible to interpret the indefinite and the name as coreferential, given a particular accenting pattern in which the name is given almost no prominence. If **Graham** in this sentence is accented, the reading is not available. If we apply the What? test (in this case the Who? test) to sentence 3.88, with **Graham** de-accented, we find, as was the case with anaphoric definites and pronouns that the sentence fails the test.

3.89 A: I met a friend on the train. Graham was on his way to an interview.
B: Who's Graham?

Given this intonation, **Who's Graham?** can only be a request for extra information, not an on-line repair of the discourse.

We will now consider each of the remaining five types of environments.

Indirect anaphoric.

3.90. A: I did not enjoy giving birth. Murdo was over twelve pounds.
B: Who's Murdo?

Associative anaphoric.

3.91. A: I quite like my family, although Margaret can be a pain.
B: Who's Margaret?

Visible situation: (at a party)

3.92. A: Elaine is drunker than I've seen her before.
B: Who's Elaine?

Immediate situation (context as for 3.65):

3.93. A: Norma is round the back.
B: Who's Norma?

The Who? question cannot, in any of these examples, be read as a legitimate request for missing information intended to facilitate repair work on the discourse.

If we consider larger situation uses of names there is an interesting effect to be observed. We can say that it is part of our background knowledge about people that they have sets of family members, friends and acquaintances and that each of these individuals has a name. We might expect therefore that in any discourse there is a p-set of named individuals associated with each participant. This *should* permit a very general use of assertive seeming names, but this does not appear to be the case. The sentences in 3.94 and 3.96 both pass the What? test although the existential propositions they are associated with are no less stereotypical than those in 3.95 and 3.97 for which What? questions are not legitimate responses.

3.94. I went out with Sonia last night.
3.95a. The speaker knows someone called Sonia.

3.95 I have left the dog in the car.

3.95a. The speaker has a dog and a car.

3.96. I have just discovered that Graham has died.

3.96a. The speaker knew someone called Graham.

3.97. I have just discovered that the butler has died.

3.97a. The speaker had a butler.

We conclude that the uses of names appear to be constrained by a stereotype effect and that the possible range of p-sets in which the referents of names are to be located is restricted to some extent. Given these observations, there does not appear to be any justification for analysing any occurrences of names as instances of assertive EF. Instead, we will assume from now on that names are unambiguously associated with pre-established EF.

3.2.2.5. The EF of Generics.

We proposed in section 3.1.2 that some uses of bare plurals should be analysed as containing a phonologically null determiner **SM** which is equivalent to the overt determiner **sm** and can be assigned assertive EF. However, since some bare plurals can only be interpreted as instances of generic reference, for example the highlighted expression in 3.98, we proposed a second phonologically null determiner **GEN**.

3.98. **Monkeys** are always the most popular zoo animal.

We propose that generics with the phonological forms of singular definites and indefinites, for example in 3.99 and 3.100 should be analysed as semantically equivalent to these bare plurals, by assuming the overt determiners to be paraphrases of the generic determiner **GEN.**

3.99. **The monkey** is always the most popular zoo animal.

3.100. **A monkey** is always the most popular zoo animal.

It must now be determined where on the existential scale generics should be placed. Carlson (1977, 1989) argues that two types of generic constructions, singular definites and bare plurals, should be analysed as proper names for "kinds of things" since these, in contrast to singular indefinites, may occur with kind-level (e.g. 3.101) as well as object-level (e.g. 3.102) predicates.

3.101. **The Panda/Pandas/*A Panda** is almost extinct.

3.102. **The Panda/Pandas/A Panda** love(s) bamboo shoots.

If singular definite and bare plural generics really are proper names, and our decision about the place of proper names on the EF scale is correct, then we would expect them to be associated with pre-established EF, but we must provide some reason for including singular indefinites at this level. Carlson's analysis of generics allows us to do this.

We propose that all objects of any particular type must be regarded as members of at least one stereotypical set, the set of all things of that type. In other words, for any individual who knows the meaning of the word **dog** (for example), the existential proposition that the set of all dogs exists can be regarded as a part of that individual's background knowledge during any discourse in which he or she participates. A non-generic specific indefinite can be then defined as an assertion of the existence of some particular entity in that set. We can then define a generic indefinite as a reference to some stereotypical member of that set. This use is not an assertion, since the knowledge that such stereotypical members exists must be pre-established. Bare plural and singular definite generics can be interpreted either as references to the set-membership as a whole (with kind-level predicates) or as further examples of references to stereotypical set members (with object-level predicates). In each case, pre-established EF is required.

One potential objection to this analysis is that it appears that generics are regularly used to *inform* hearers of the existence of objects of some type. Sentence 3.103, for example, would be appropriate in the setting of a lecture.

3.103. **Hyphomycetes** are fungi which produces spores through pores.

However, we can argue that the set <**hyphomycete**> is not necessarily mutually accessible to the speaker and hearers of this discourse since an utterance of 3.103 does not constitute an act of generic reference to the set or to stereotypical members of the set, but instead an initial act of naming in the sense of Kripke (1980) which introduces the set to the hearers. If we can analyse such acts of naming as distinct from acts of generic reference, the way is clear to claim pre-established EF for all instance of generic reference.

3.2.2.6. The EF of Every.

In semantic theories, such as Generalised Quantifier Theory (Barwise and Cooper, 1984, Partee, ter Meulen and Wall, 1990) which view existential import as a binary logical feature, any quantifier or determiner which does not entail existence has no EF. However, we will argue that **every**, for example, should be associated with a level of EF distinct from pre-established or assertive EF, but distinct also from the total absence of EF associated with the quantifier **no**. The examples in 3.104 and 3.105 show two distinct readings of **every** an existential reading in 3.104 and a non-existential reading in 3.105.

3.104. I thanked **everyone who helped**. **They** had been very kind.

3.105 speaker A: So why didn't you write to me?
speaker B: You didn't write to me!
speaker A: I answered **every letter I got**.

To account for these two readings, we could say that **every** has no existential force and that the proposition (in 3.104) that there exists some individual who helped comes about as a result of the use of the pronoun **they**. However, it was argued in section 3.2.1.2 that a pronoun has pre-established EF, i.e. that the proper use of a pronoun requires that the existence of its referent is already established. The cross-sentential anaphors of **every,** if it was analysed as non-existential, would contradict our claims regarding this feature of the EF of pronouns.

Kamp and Reyle (1993) discuss the cross-sentential anaphoric potential of **every** at some length and propose an operation which they call *abstraction,* the function of which is to provide **every** with a discourse referent which can then be linked directly to the anaphor by co-indexing. This type of treatment accounts for the facts, but again the abstraction operation is triggered by the occurrence of the anaphoric pronoun rather than by the occurrence of the antecedent containing **every**, and so it appears that under their treatment too, a requirement that the existence of an anaphoric pronoun be pre-established is not met.

It appears that if the treatment of **every** as an antecedent is not to interfere with the general requirement that the existence of referents for pronouns is pre-established, then **every** must be analysed as unequivocally associated with positive EF. However, the impossibility of *this* analysis is shown by the short discourse in 3.105 repeated here, where it is possible for all three sentences to be true.

3.105 speaker A: So why didn't you write to me?
speaker B: You didn't write to me!
speaker A: I answered **every letter I got**.

The truth of speaker A's second utterance rests on the possibility of a non-existential reading **every** allowing the interpretation of the utterance as an expression of the proposition that since she received no letters from B, it is true that there is no letter which she received but did not answer. Non-existential readings for every are not common, but they do occur. Most naturally occurring instances from real discourse are jokes, such as 3.106:

3.106. Bernard Manning is not offensive to women. **Every feminist with a sense of humour** adores him.[13]

Even though non-existential readings are stylistically marked, because the sentences associated with them are true, we cannot simply disregard these instances as mis-uses. There is further motivation for a treatment of **every** which allows both readings when we consider a problem identified by Fodor (1979) in her account of the asymmetry of opaque predicates.

Fodor characterised the asymmetry between the affectedness of the arguments of an opaque predicate such as **think of** by stating that, given two individuals, Xenon and Yolande, of which only Xenon is accessible, there is "nothing we can do" to Xenon to bring about a state of affairs about which sentence 3.107. is a true statement.

3.107. Yolande is thinking about Xenon.

If we have access to Yolande alone, on the other hand, Fodor observes that we can bring about the state of affairs which 3.108 describes.

3.108. Yolande is thinking about everyone who is eating cheese.

The problem Fodor identifies with this analysis is that:

> "given that Yolande is thinking about everyone who is eating cheese, then by getting Xenon to eat cheese it would appear that we could make it true that Yolande is thinking about Xenon."
>
> (Fodor, 1979: 215)

Given an analysis of **every,** which allows both existential and non-existential readings, the solution to this problem is easily identified. A sentence such as 3.108 is true if and only if every member of a set of people who are eating cheese is accessible to Yolande. In other

---

[13]The semantic plurality of **every** which contrasts with its syntactic single number can also be cancelled and this also appears usually to be feature of jokes. For example: **Every Tory voter in Scotland has turned out to greet you, Prime Minister. Shall I show him in?**

words, the possibility of this statement being true rests on the possibility of an existential reading of **every** which facilitates a *de re* interpretation of **everyone who is eating cheese**. If we add Xenon to the set, the set membership changes, it may no longer be the case that Yolande has access to the entire set of people who are eating cheese, and so the statement may no longer be true. Sentences such as 3.109, however, do not show this effect.

3.109. Yolande, a vegan, would like to kill everyone who is eating cheese.

The truth of 3.109 does not require an existential reading of **every** and in these cases, on a *de dicto* reading of **everyone who is eating cheese,** we can add Xenon to the set of cheese-eaters without causing a change in the truth-value of the sentence.

The preceding discussion indicates that the most satisfactory analysis of **every** is one which accounts for its cross-sentential anaphoric potential *and* allows for the infrequent instances where it must be understood as non-existential. We conclude that the EF of every should be analysed as a strong, but cancellable, implicature.

Implicature as characterised by Grice (Grice 1975) is a scale of declining strength with three slots; conventional (for example the temporal ordering conveyed by **and**), generalised conversational (for example, the causality conveyed by **therefore**) and particularised conversational implicatures. Since we are attempting to show that existential force is also best seen as a scale of declining strength, it is worth looking to see if the various determiners which are not associated with pre-established or assertive EF, could be slotted into the spaces on the Gricean scale with respect to their existential import.

Grice defined typical conventional implicatures as determinate and uncancellable, in contrast with generalised conversational implicatures which are typically indeterminate and cancellable. The existential implicature associated with **every**, however, is determinate in that it always concerns the existence of some entity in the denotation of the common noun it quantifies, but is cancellable. It is immediately clear that existential implicatures will not fit neatly into the Gricean scale. We will not therefore, pursue the classification of this and the weaker levels of EF using the standard characteristics of conventional or conversational

implicatures, but will simply say that **every** is associated with the strongest possible cancellable implicature of existence, which allows subsequent pronominal reference to the object whose existence it suggests (as in 3.104) and that cancellation of the existential implicature is stylistically marked.

We cannot end a discussion of the existential readings and anaphoric potential of **every** without considering the issue of donkey anaphora. Sentences such as 3.110 have been notoriously problematic for analyses which translate **every** into the universal quantifier of classical predicate logic since in the logical translation the last x and y variables are outside the scope of the quantifier and yet the interpretation of the sentence demands that the pronouns are bound by the quantifier, since the reference of **he** and **it** vary according to the reference of **a farmer** and **a donkey**.

3.110. Every farmer who owns a donkey beats it.

3.110a. $\forall x$ (farmer (x) & $\exists y$(donkey(y)) & owns (x,y)) beats (x,y)

The Discourse Representation Theory approach to quantification and anaphora has solved the problem of the translation of donkey sentences into predicate logic formulae. We cannot go into the details of the DRT approach here, but very briefly, DRT introduces an intermediate stage (the Discourse Representation Structure) between the sentence in natural language and the logical translation. At this stage **every** is represented as a conditional and new variables are introduced which serve as "discourse referents" for the pronouns. This DRS is then translated into predicate logic as:

3.110b. $\forall xy$ ((farmer(x) & donkey (y) & owns (x,y) $\rightarrow$ $\exists uv$ (u = x & v = y & beats (u,v)))

On the approach to the interpretation of **every** taken in the current thesis, donkey anaphora is not a particularly problematic issue. It appears that the interpretation of **every** in donkey sentences, as in other contexts, can be either existential or non-existential, since there are

acceptable continuations which both uphold and cancel the existential implicature. For example:

3.110. Every farmer who owns a donkey beats it.

a. Thankfully The Sun has adopted all of them and put them out to grass at Southend.

b. Their suffering is indescribable.

Whether **every** is interpreted existentially or non-existentially, it seems the sentences are best viewed as variants of particularly strong generics; they can be paraphrased by any of the following:

3.112. If a farmer owns a donkey, he or she beats it.

3.113. Farmers who own donkeys beat them.

3.114. The farmer who owns a donkey beats it.

In any case the problem of the fixing referents for **he** and **it** does not arise if we accept the broad characterisation of anaphora proposed by Hawkins (1978) and discussed in section 3.2.2.2. above.

3.2.2.7. The EF of Any and Few.

We would expect to find, in expressions with weaker existential implicatures than that associated with **every**, either less potential for the expression to appear in antecedents of cross-sentential anaphoric pronouns, or that the cancellation of the existential implicature would be less stylistically marked, or both.

There does not appear to be any strong evidence that **any** and **few** are less able than **every** to function as cross-sentential antecedents. The following three examples seem equally acceptable.

3.115. I am determined to kill every rabbit living in my garden. They have eaten all my seedlings two years in a row.

3.116. I am determined to kill any rabbit living in my garden. They are eating everything I plant within a day.

3.117. So far I have witnessed very few rabbits in the act, but at least I have been able to kill them[14].

However, there does appear to be evidence that the existential propositions associated with **any** and **few** are more readily cancellable than those associated with **every** (e.g. 3.118 compared with 3.119) and that their cancellation does not have so marked a stylistic effect (e.g. 3.120 compared with 3.121.).

3.118. ??I expect I will be bitten by **every hamster in the show**, but luckily only dogs and cats are entered.

3.119. I am free to discuss my case with **any doctor in the hospital**, which would be more of a comfort if the Trust could afford to have doctors on the staff.

3.120. ??I plan to dance with **every woman in the room**, and since it is an all-male gathering that will leave plenty of time for cards.

---

[14]In addition to this type of anaphoric link, **few** can also occur as an antecedent for what we might call complementary anaphors, such as them in sentence 3.113.

3.113. So far I have managed to kill very few rabbits, but I am determined that I will get **them** in the end.

The referent of the highlighted pronoun in 3.113 is the complement of the sub-set of rabbits picked out by the expression **very few rabbits**. This referent is located by subtracting the referent of **very few rabbits** from the p-set of rabbits in the discourse context. It is a general feature of monotone decreasing determiners such as **few**, **less**, and **no** that they can serve as antecedents for complementary anaphors.

3.121. I will happily promise to kiss **any frog in the pond** since I know there are only goldfish and ducks.

We propose, on the strength of these of examples, that the EF of the quantifiers **any** and **few** should be analysed as a level of implicature weaker than that associated with the EF of **every**.

Before we leave the discussion of the EF of **any**, we will consider the question of whether there is any support from existential aspects of the meaning of **any**, for an analysis of this quantifier as ambiguous between two readings: **any**$_1$ which is polarity sensitive (e.g. 3.122) and **any**$_2$ which has been called "free context **any**" (e.g. 3.123) (Ladusaw 1979).

3.122. I have never seen **any**$_1$ **dolphins** in Scottish waters.

3.123. **Any**$_2$ **cabby** will let you pay by cheque if you tip in cash.

The label "free-context any" is somewhat misleading, since the contexts, besides negatives, in which **any** may occur are highly constrained; being restricted to modal contexts (e.g. 3.123), antecedents of conditionals (e.g. 3.124), and generics (e.g. 3.125).

3.124. If I hear **anything**, I'll ring you.

3.125. I like **any** food with lots of calories.

There does not appear to be any reason to analyse **any** as ambiguous with respect to the strength of its EF since if we simply extend the set of permissible contexts for the single unambiguous **any** to encompass both so-called free-context positions and positions within the scope of a negation operator we can say that in each environment the existential proposition can either be cancelled or upheld and that the resulting truth-value is predictable. For example, 3.126 and 3.127 are acceptable on either the a or the b readings.

3.126. **Any** dragon would nip you over the channel for a fiver.

3.126a. And it is usually easy to find **one**.

3.126b. But since there are **none** you will have to pay the ferry fare.

3.127. I have never seen **any** mermaids in Scottish waters.

3.127a. But I know that there are lots of **them**.

3.127b. I am not too hopeful of doing so since there are **none**.

Unlike the continuations of statements with expressions containing **every** however, the b versions of 3.126 and 3.127 which cancel the existential implicature of **any** seem marginally less marked than the a versions which uphold it. Furthermore, the most acceptable continuations tend to be ones in which the pronoun which picks out the referent of the expression containing **any**, is not in a completely transparent context. Both of these observations offer further corroboration of our analysis of the EF of **any** as an extremely weak implicature.

3.2.2.8. The EF of **no**.

The lowest level on the scale of EF is that associated with the determiner **no** and its pronominal variant **none**, e.g. the object NPs in 3.128 and the second conjunct of 3.129.

3.128. The pet shops around here have no bilingual hamsters.

3.129. I wanted a bilingual hamster but the pet shop had none.

The EF of **no** and **none** is straightforward to classify since the function of **no** in some NP is to assert the non-existence in the context of any entities which satisfy the descriptive predicate of the NP. The precise details of the interpretation of determiners such as **no** and **none** have a attracted a degree of interest in formal approaches to quantification which analyse **no** as containing a quantifier and a negation operator (May, 1985). The focus of such work has mainly been to provide a translation into a logical language in which the relative scope of the quantifier and the negation operator reflects the interpretation of the corresponding sentence in natural language. For example, such theories seek to reflect the truth-conditional non-equivalence of sets of sentences such as 3.130. - 3.132.

3.130. I did not tell everyone.

3.131. I did not tell anyone.

3.132. I told no one.

With respect to the purely existential features of **no** and **none**, there may be semantic effects to be observed relating to a possible distinction between references to backgrounded propositions of non-existence and new assertions of non-existence, but this potential area of research lies beyond the scope of the present study. We conclude simply that expressions containing the determiners **no** and **none** are incompatible with the existence of any referent of the type denoted by the descriptive predicate.

3.3. Conclusion.

In this chapter we have concentrated on the existential possibilities associated with linguistic expressions. We have argued that to account for the pattern of occurrences of NPs of different types, and the anaphoric potential of these NPs we must recognise that existential force is a scalar phenomenon comprising five levels and that NP types should be placed on the existential scale in the order in which we have discussed them above: names, pronouns, definites and generics at the highest level, followed by indefinites, **every** N, **any** N and **few** N, and **no** N. In the next chapter we turn our attention to the problem of the existential possibilities associated with entities. We will formulate a theory of possible worlds which allows us to account for the types of existence which hold of entities, as opposed to linguistic categories, and in the following chapter, we combine both types of existential features to develop an account of the interpretation of the expressions we have been examining in real discourse.

Chapter 4.
Existential Features of Entities.

4.1. Types of Existence.

There is a long and prolific body of work in the philosophical discipline concerning questions of existence[1] which the current thesis, to a large extent, ignores. The justification for our decision not to discuss the philosophical work on existence in any detail must be that our interest in the questions surrounding this area is of a very limited and specific kind. We are seeking to account for the effect that the existence and non-existence of objects has on the *use* of referring expressions as attempts to pick out these objects. Our first concern is to develop whatever approach to the existence of entities and the status of possible worlds will allow an accurate and *linguistically* well-motivated characterisation of reference and truth-value assignment. Although we would not posit any view of the existence of entities which flies in the face of established philosophical views, without some compelling motivation, our aim is to construct a linguistic, not a philosophical theory.

To return to the question of types of existence, we could predict that a philosophically naive informant, if asked to decide how many kinds of existence there were, by making a decision about how many occupiable slots there were on a scale between existence and non-existence, would reply that there were two; that either something exists (e.g. the King of Jordan) or it does not exist (e.g. the Queen's second daughter); that there can be no middle ground. However, although this limitation is initially appealing on an intuitive level for such clear cut-cases, it appears less useful when we consider the list of entities in 4.01.

4.01. John Major, Pitt the Younger, Santa Claus,
Bilbo Baggins, Chelsea Clinton's sister.

It is to be expected that there would be considerable disagreement in any attempt to distribute these individuals between the two slots. It would be reasonable to claim that only John Major

[1] for example, Carnap (1947), Strawson (1950), Quine (1953), Geach (1959), Kripke (1980)

exists, but it would be equally reasonable to claim of the non-existent entities that they each fail to exist in distinct ways. We could say that Pitt the Younger does not exist any more; Santa Claus does not exist outside European folklore; Bilbo Baggins does not exist outside Middle Earth; and Chelsea Clinton's sister does not exist at all. So might this mean that we must, after all, recognise a variety of levels of existence? If we want to avoid such a step and instead retain the binary logical definition of existence, we could reason that each of the "non-existent" entities can be characterised by a distinct exception to its non-existence.

But if we accept that there are exceptions to the non-existence of the "non-existent entities", it is unclear on what grounds we could continue to define them as non-existent. This becomes still harder to defend when we reflect that the entity (John Major) which we defined as having existence in an ideal sense, itself fails to exist in some of the other locations. For example, John Major does not exist in European Folklore or in various fictional locations such as Middle Earth. It is more straightforward, in view of this observation, to say that John Major, Santa Claus and Bilbo Baggins all exist, but in different locations. This allows us to continue to hold the concept of existence as an essentially binary property, since any entity either exists or does not exist at some location. Furthermore, if we are prepared to accept the idea, generally assumed in formal semantics, that the same spatial location at two different times constitutes two separate locations i.e. that the slots in which entities occur are best characterised as indices which are space-time pairs, then we can also claim that Pitt the Younger and John Major exist at different locations.

On the other hand, when we say that Chelsea Clinton's sister does not exist at all, what we are saying is that there are no *identifiable* exceptions to her non-existence, no easily named or recognised location in which she might exist. We are not saying that there is no *conceivable*, *potential* exception to her non-existence. We are happy to accept that, in some location, an individual might have existed or might yet exist who could be referred to as **Chelsea Clinton's sister.** On the basis of this, we propose a further type of existential location, which can be called a possible (although non-actual) location and we are now able to claim that all of the entities in 4.01 exist, in a range of locations.

The remainder of this chapter and the next chapter are taken up with a discussion of one particular view of what might constitute the set of actual and potential existential locations; that is, the notion of a set of possible worlds.

4.2. Possible Worlds.

The way in which possible worlds are put to use in the current thesis differs considerably both from their original purpose in modal logic and from their place in model theoretic formal semantics. The notion of possible worlds was first used in the field of logical linguistics in the analysis of the semantics of alethic modal statements (Lyons, 1977). Necessary truth had been defined by Leibniz as truth in all possible worlds and contingent truth, as truth in at least one possible world. From this basis, a characterisation of all of epistemic and deontic modality in terms of possible worlds could be derived. Possible worlds used for this purpose can be thought of as alternative (non-anomalous) states-of-affairs (McCawley, 1981) or simply "ways in which the world could have been[2]" (Allwood, Andersson and Dahl, 1977). In Montague Grammar, a set of possible worlds is included as one of the primitives of the model with respect to which *any* sentence, not only sentences expressing modal statements, is evaluated. Here, the set of possible worlds has identical ontological status to the set of primitive entities, the set of primitive times and the denotation assignment function. The truth-conditional meaning of a declarative sentence, in Montague Grammar, is therefore taken to be the truth-value of that sentence with respect to some model which specifies a sub-set of entities, a denotation assignment function, a particular world, and a particular time (Cann, 1993; Dowty, Wall and Peters, 1981).

In short, the employment of an infinite set of possible worlds is now unremarkable in linguistic semantics[3] as well as in the fields of logic and philosophy. However, although possible worlds are in general use as a theoretical tool, there is a considerable amount of disagreement surrounding them. There are several distinct and contradictory definitions of possible worlds

---

[2]or "must have been" or in deontic modal statements "ways in which the world is obliged/permitted to be" or in epistemic modal statements "ways in which the world is believed/known to be".
[3] see for example Bradley and Swartz (1979); McCawley (1981); Green (1989); Cann (1993)

and little consensus concerning what exactly they are. Two main classes of definition stand out.

According to one type of definition, a possible world is a type of entity. This definition derives originally from Wittgenstein's definition of a world. To define the real world, Wittgenstein proposed an ontology which consisted of four primitives: objects and individuals, which together make up the class of items, and properties and relations, which together make up the class of attributes. He proposed that items and attributes together make up states of affairs and then defined the real world as "the totality of actually existing states of affairs" (Wittgenstein, 1921). By analogy with this definition of the real world, we can give a definition of a possible world as an entity which is the totality of some set of possibly existing states of affairs .

Other logicians, philosophers and linguists have taken possible worlds to be abstract constructs. For example, for McCawley, a possible world is an abstract linguistic construct defined as "any assignment of the values true and false to all propositions of a given language"[4]. An impossible world is defined as one which does not conform with the rules of inference in a given system (McCawley, 1981). Similarly, Allwood, Andersson, and Dahl define a possible world as "the set of propositions which are true in it (and thus describe it)." (Allwood, Anderson, and Dahl, 1977).

In contrast to both of these stances, some linguists acknowledge the philosophical problems which seem to be inherent in the concept of possible worlds but explicitly set aside the issue and rely on a functionally and technically adequate definition. For example Cann (1993), makes no decision about the status of possible worlds but *treats* them as abstract constructs:

---

[4]McCawley actually uses the term "non-anomalous state of affairs" but explicitly states that the term is intended to be synonymous with "possible world".

> "...the adoption of a set of primitive worlds is somewhat mysterious ... the ontological status of possible worlds has been the subject of much philosophical debate, but it is not my intention to go into this. If we assume that possible worlds are theoretical constructs (whose ontological status is similar to other mathematical constructs) then what we as linguists need to do is assess their usefulness in achieving the general aims of semantic theory..."
>
> (Cann, 1993: p270)

Other linguists appear to attempt to reap the benefits of a particular view of the nature of possible worlds while leaving an escape route to be used if the holding of that particular view entangles them in philosophical difficulties. For example, Fodor (1979) would like to claim that possible worlds are more like real entities than abstract constructs but in the end ducks the argument:

> "... I shall assume an ontology of possible worlds ... The "psychologically real" ontology is the one relevant to explaining the semantics of a natural language, even if upon examination it proves indefensible as a theory of what "really exists" in our universe. But in any case, as far as I can determine the references in what follows to non-real worlds and individuals are innocent and could be translated by anyone who is squeamish about them into less loaded (though also less vivid) terms."
>
> (Fodor, 1979: p2)

Not all philosophers take any more seriously than linguists the problem of the actual nature of possible worlds. For example, the philosophers Robert Bradley and Norman Swartz in their text book on possible worlds assume with little preliminary argument that possible worlds are abstract entities theoretically equivalent to such other abstract entities as numbers (Bradley and Schwarz, 1979). However, Lewis (1986) expounds a theory of possible worlds, which he calls *modal realism*, that consists in a kind of super-real definition of a possible world. He holds the view that all possible worlds actually exist, i.e. that they have the same existential status as the real world, and argues at length against approaches which allow possible worlds into their ontology as abstract entities; labelling the resulting theories scathingly as examples of various kinds of "ersatzism" and claiming that definitions such as McCawley's (linguistic ersatzism) are no more than attempts to gain "paradise on the cheap"; by which we can take it that he means to suggest that they are unacceptable attempts to exploit the power which

possible worlds contribute to a theory without having to countenance what even *he* admits is the counter-intuitive notion that all possible worlds "actually" exist.

However, given that our interest is specifically linguistic, we are inclined to give more weight than Lewis to linguistic arguments in favour of one or the other type of definition. In fact, even from some linguistic points of view there is conflicting motivation. It is problematical to argue from a purely formal linguistic framework that worlds are in any way distinct from their descriptions since the utility of possible worlds is bound up with, and inseparable from, the characteristic property that two worlds are distinct if and only if there is some statement in a language that is true in one world but false in the other. If possible worlds are analysed as existent entities, then two worlds may be distinct even if for every statement in a language the truth-value of that statement in one world is matched by its truth-value in the other.

On the other hand, if possible worlds are analysed as existent entities the linguistic problem of characterising translation is simplified, since we can say that given two languages L1 and L2 and two statements ϕ and ψ in L1 and L2 respectively, ϕ can be defined as a translation of ψ if and only if the set of worlds in which ϕ is true is identical to the set of worlds in which ψ is true[5]. If possible worlds are not analysed as existent entities but as partitions on a vocabulary, then it becomes impossible to define translatability since the set of possible worlds described by L1 is of necessity distinct from that described by L2.

Despite such difficulties, there are several reasons why, in the context of the current thesis, it is not satisfactory to leave the problem of the essential nature of possible worlds on one side and simply use the notion as a theoretical tool. We aim to define each possible world in terms of its accessibility from, and proximity to, other worlds and in this way to build an extremely structure-rich arrangement of worlds in which each individual world is uniquely determined along a number of mutually independent parameters[6]. In the pursuit of this aim we have to provide answers to a number of specific questions concerning properties of possible worlds

---

[5]and this can be parameterised to handle near misses.
[6]see, for example, chapter 5, section 5.1.1

such as whether it is possible for worlds to be underspecified[7] or whether there is any such thing as a impossible world.

Since the features of possible worlds are to be defined relationally, our task is only possible if we can start from one world which has its nature and features independently fixed and define other worlds comparative to it. The world which appears to be the best candidate to be defined external to the theory is the real world. If we accept that the real world is the most likely world to be an existent entity and we agree to accept our intuitive judgement about what the real world is, then by deciding to view all other possible worlds as essentially the same as the real world (i.e. as existent entities) we are able to make predictions about the properties of worlds about which we have no direct knowledge. There is in addition some specifically linguistic motivation for analysing possible non-actual worlds as equivalent to the real world in that there is no evidence from the comparison of the behaviour of the real world with that of non-real worlds in discourse which convinces us of the necessity to treat them as essentially different. In fact, there are several potential problems involving reference and truth-value assignment which are most easily handled if all worlds, real and non-real, are taken to be of the same kind.

However, one problem with the modal realistic view is that it is difficult to give a definition of the primitive entities of which worlds consist. If we return to Wittgenstein's definition of a world as a totality of states of affairs made up of items and attributes, it is easy to see that, on a modal realistic approach, we can define in a straightforward way what the set of items in a world contains: concrete objects such as tables, squirrels and Zsa-Zsa Gabor; groups of concrete objects such as species, generations and forests; concrete masses such as oceans, deserts and wine. It is hardly more problematic to accept the existence of abstract objects such as ideas and numbers; groups of abstract objects such as manifestos; and abstract masses such as integrity and love. However, it is difficult to define the nature of the other type of primitive entity; the set of attributes.

---

[7]c.f. chapter 5, section 5.3.1

If a world is an abstract linguistic construct then the set of attributes is defined in a straightforward way by the denotation assignment function, but if worlds exist independently of any language then properties and relations cannot be defined in this way. We are forced simply to accept the notion that worlds contain two types of primitive entities: items - things in the world, and attributes; ways that things are. Note that we define attributes as ways that *things* in worlds are and not ways that worlds are. This reflects our claim that sentences are about individuals in worlds not about worlds themselves and that statements are true of entities in worlds not true of worlds themselves.

4.3 The Choice of Reference World.

Clearly a participant in some discourse can refer to and make statements about things which do not exist in the world in which the discourse takes place. For example, Sentences 4.02 - 4.05 can be read as tokens of true statements about entities which exist in four worlds which are, in various ways, distinct from the world in which this thesis is being written.

4.02. The sun never sets on the British Empire.

4.03. Sherlock Holmes puffed on his pipe and pondered.

4.04. Bilbo baked bread.

4.05. The Tooth Fairy had a busy night.

The British Empire picks out an entity which exists in the same world at an earlier stage; **Sherlock Holmes** picks out an entity in the fictional counterpart of the real world described in the novels of Arthur Conan Doyle; **Bilbo** picks out an entity which exists in the fictional world Middle Earth, and **the Tooth Fairy** picks out an entity which exists in what we could call a mythological counterpart of the real world. As well as such earlier, fictional or mythological worlds, discourse participants are also able to invoke an infinite variety of other worlds "off the cuff". For example, we would expect that first-time hearers or readers of 4.06 would be unable to interpret it in any other way than as a factual statement about an entirely unfamiliar world.

4.06. **Chelsea Clinton's sister** had a tummy tuck.

In addition, speakers can also invoke what we could call temporary counterparts or alternatives to the real world. The sentences in 4.07 and 4.08, for example, might be read not as factual statements about non-real worlds but as predictions about alternatives to the real world.

4.07. Chelsea Clinton's sister might be called Belgravia.

4.08. My daughter would never play with Power Rangers[8].

Having proposed the inclusion in our ontology of a set of possible worlds which include such unfamiliar worlds as that in which Chelsea Clinton's sister exists, there is no rationale for ruling out any world in which Chelsea Clinton's sister has, or does not have, any particular property. If we accept, for example, the existence of a world in which Chelsea Clinton's sister exists we must inevitably recognise the existence of a world in which Chelsea Clinton's sister has a tummy tuck. In other words the potential truth of sentence 4.09 entails the truth of sentence 4.10.

4.09. Chelsea Clinton's sister had a tummy tuck.

4.10. There is a world in which Chelsea Clinton's sister had a tummy tuck.

If we analyse every referring expression as picking out an entity in some possible world, we are forced to accept that, for any sentence we utter, there is some world in which it is true (and some world in which it is false). But we are *not* proposing that the role of possible worlds in the process of interpretation and truth-value assignment is to facilitate the analysis of sentence tokens as true statements, since recognising that there must be some world in which a sentence is true, is clearly distinct from *judging* a particular *utterance* of that sentence to be true. In other words, any particular token of sentence 4.09 is not inevitably perceived as true, on the basis that sentence 4.10 is necessarily true. If we believe that any sentence that is not

[8]In a context in which the speaker has no children

anomalous[9] is true in some possible world, and if we were to accept 4.10 as an accurate *paraphrase* of 4.09, rather than merely an entailment, it would follow that any utterance of any non-anomalous sentence would be trivially true, and hence massively uninformative. We claim instead that a knowledge of truth-conditions is not part of interpretation, but rather a *prerequisite* for successful interpretation. Knowledge of lexical meaning and syntax between them constitute the potential ability to state the truth-conditions of any of the infinite number of possible sentences of a language. But this does not enable any language user to assign a truth-value to so much as one utterance of any of these sentences. We could say that knowing what a sentence means - knowing its truth-conditions - relies on our (possibly unconscious) acceptance that a sentence is true in some possible world. However, this knowledge alone never leads to an assignment of a truth-value of true. Likewise, knowing a sentence's truth-conditions entails that we accept that there is (at least) one world where that sentence is false but this does not lead us to interpret any utterance of that sentence as false.

The process of interpretation may consist of many more tasks than truth-value assignment although it is our contention that truth-value assignment is an essential part of the interpretation of the utterance of any proposition. The simplest type of interpretation process therefore consists of truth-value assignment alone. This is the case when the intended communicative effect of some utterance is nothing more than the transference of the information contained in the proposition lexically encoded by the sentence. However, even in this simplest of cases, it is rarely the hearer's aim to assign the truth-value *true* to an utterance. The hearer's aim is to assign the appropriate truth-value according to whether or not the utterance reflects a state of affairs in an accurate way, and since truth-values are defined with respect to particular referents, and referents are to be located in any of a variety of worlds, we can say that an important initial step in the process of truth-value assignment is the choice of reference world.

The entities with respect to which the speaker intends some statement to be evaluated must in every case (including all of those in 4.06 - 4.10 above) be inferred by the hearer. That is, excepting the referents of deictic expressions (although even in these cases the hearer must

---

[9]In chapter 5 we discuss the question of anomaly when we consider whether there are such things as impossible worlds.

decide that they *are* deictic) the referent has to be located by the hearer. To say that the hearer must locate the referent, having claimed that the hearer must recognise that potential referents exist in an infinite number of distinct possible worlds, is equivalent to saying that the hearer must choose which possible world is the correct location for the referent of an NP in any particular case.

The choice of reference world, as part of the process of interpretation, could be seen as the result of a decision about how far a hearer will go towards the possible world where any sentence is true. The important task for this thesis is to identify the rules which govern the amount and kind of world-shifting which goes on during interpretation.

To determine whether the rules we propose are correct, we will rely on native speaker intuition and to harness our intuitions it is necessary to disregard some kinds of discourse contexts since they are associated with fixed conventions with regard to how much world-shifting is permitted to take place during the interpretation of utterances which occur in them. Contexts in which unconstrained world-shifting takes place include story-telling discourses. Contexts where any world shifting is prohibited include the giving of courtroom evidence, where not only are false factual statements about the real world viewed as such, rather than as true factual statements about some other world, but modal statements are constrained and it is not permissible to state propositions which rely on certain types of knowledge, such as hearsay.

However, courtroom evidence apart, the intuition is that it is rarely the case that the real world has the kind of exceptional status in terms of truth-value assignment which is sometimes claimed for it (e.g. in Russell's and Strawson's approaches to reference and truth-value assignment). We claim that hearers do *not* generally interpret as true only those sentences which are true of entities in the real world. Nor are true-in-the-real-world sentences interpreted as in some way more true than other sentences. For example, we claim that sentences 4.11 and 4.12 are equally open to an interpretation which assigns to them a truth-value of *true* even though the named entity in 4.11, Mr Major, is an inhabitant of the current world, while the named entity in 4.12, Santa Claus, is existentially more questionable.

4.11. Mr Major wears a grey suit.

4.12. Santa Claus wears a red suit.

We might want to say that Santa Claus exists in the current world in that representations of him exist in the current world and in the overwhelming majority of these he is depicted wearing a red suit. We could then be forced to admit that both 4.11 and 4.12 are true statements about entities which exist in the current world. If indeed **Santa Claus** is the name of a set of entities of a recognised type which exist in the current world we can explain why sentence 4.12a. is perceived to be true and roughly equivalent to sentence 4.12.

4.12a. Santa Clauses wear red suits.

However, we can identify properties which hold of Santa Claus, the entity which exists in mythology, but not of the Santa Clauses, i.e. the representation of Santa Claus, pictorial or animated, which exist in the real world. 4.12b. and 4.12c. show examples of these:

4.12b. Santa Claus comes down chimneys in the middle of the night.

4.12c. Santa Clauses come down chimneys in the middle of the night.

For 4.12b. to be perceived as true it must be a statement about the mythological entity rather than the set of real world entities; it cannot be paraphrased as 4.12c. without a change in truth-value to *false*. Since 4.12b. is indeed perceived as a true statement we can conclude that the name **Santa Claus** in it picks out a entity which does not exist in the current world.

To summarise, we can say that to know what a sentence means, to know its truth-conditions, is to recognise the necessary properties of the possible worlds in which it is true and the necessary properties of the possible worlds in which it is false, and this in turn requires the implicit acceptance that there is some possible world in which it is true and some possible world in which it is false. However, there is no motivation in the process of interpretation and truth-value assignment to take the step from the knowledge that an utterance represents a sentence that is true in some possible world to the judgement that the same utterance expresses

a true statement with respect to a specific world; i.e. that it will be perceived as true once the reference world has been chosen. During truth-value assignment, the identity of the world or worlds in which referents are to be located, far from being given in advance, has to be fixed as part of the process of interpretation. We will call this view of the process of the interpretation of referring expressions, which states that a referent must be located in one of a set of possible worlds in this way, *a location theory of reference*.

The term "location" has been used before in connection with questions of reference, in Hawkins (1978; 1984). Hawkins claims that some uses of definites and indefinites instruct the hearer to locate a referent in a p-set, a shared set of entities (Hawkins, 1978: 167). For example, the use of the definite in 4.13, on Hawkins theory, would constitute an instruction to locate the referent of **the class** in a p-set (e.g. to understand it as reference to a class that the speaker is a member of, or is the teacher of etc.). The indefinite then instructs the hearer to locate the referent of **a student** in the p-set invoked by the utterance of **the class.**

4.13. The class broke up early because a student set off a stink-bomb.

In the terms of the theory developed in this thesis, we agree with Hawkins that the location of the referent of **a student** is governed by a relation of associative anaphora with **the class;** the question relating to location which concerns us is what governs the choice between locating the referent of **the class** in the real world or some other possible world.

4.4. Initial Motivation for Location in Possible Worlds.

Our theory of the interpretation of referring expressions, which stipulates that all referring expressions must have referents, since they are interpreted by a process of locating those referents, is obviously at odds with the view that potentially referring expressions do not have referents in particular environments, known as opaque contexts. It is also incompatible with the view that truth-valuelessness can be explained as a result of reference failure. An in-depth discussion of the efficacy of a location theory using possible worlds in accounting for patterns

of truth-values is not given until the next chapter. However, before setting out a detailed account of the structure of the set of possible worlds, we will offer some initial justification for the theory by comparing the current approach with two others and will show that neither an analysis of "referring expressions" without referents in opaque contexts, nor an analysis which rests on the notion of reference failure, can account for the full range of data.

4.4.1. Location and Opaqueness

It has been argued that nominal expressions in opaque positions do not refer (Quine, 1960). We will show that such an analysis is flawed. To do this we will look in some detail at the behaviour of such expressions in what are called opaque predicate positions; i.e. those contexts in which the Law of Substitution fails. Examples of verbs which express opaque predicates are **dream about, admire, hate, worship,** and **want**. In the current context, the important property of an opaque environment is the property which has been defined and explained away as an ability to cancel the existential presupposition of its object NP. For example, the object NP in 4.14 would not be analysed as having an existential presupposition in those cases where the verb expresses an opaque predicate (the examples without asterisks).

4.14. I { *met / admire / dreamed about / worship / *bit } the Queen's second daughter.

Frege accounted for this by arguing that in such contexts expressions denote their intensions (Sinnen), while they refer to their extensions (Bedeutungen) in transparent contexts (Geach and Black, 1980). On this analysis, since normally referring expressions do *not* refer, it is not sensible to talk of any requirements of existence on their referents. For example, 4.17 and 4.16, as well as 4.15 have the potential to be interpreted as true statements, in the current

context where the speaker/writer is the writer of this thesis, if **dream about** is analysed as an opaque predicate[10].

4.15. I dreamed about John Major's brother last night.

4.16. I dreamed about Sherlock Holmes' brother last night.

4.17. I dreamed about Chelsea Clinton's sister last night.

What we will call **the opaqueness explanation** begins with the premise that failed existential presuppositions (themselves brought about by the non-existence of a referent) result in the truth-valuelessness of the presupposing sentence. To account for the possibility of a truth-value being assigned to sentences such as 4.15 - 4.17 the opaqueness explanation appeals to the notion that the existential status of the referent of the NP in object position is irrelevant to the process of truth-value assignment. The opaqueness explanation would predict truth-valuelessness for sentence 4.18 since **Chelsea Clinton's sister** occurs in a transparent, presupposition preserving position and its referent does not exist, and indeed this accords with our inability on an intuitive level to assign a truth-value to this sentence.

4.18. Chelsea Clinton's sister dreamed about Sherlock Holmes.

However, the opaqueness explanation *cannot* account for the perceived truth-value of 4.19.

4.19. Sherlock Holmes dreamed about Chelsea Clinton's sister.

The opaqueness explanation would predict either that, if Sherlock Holmes is taken to be an existent entity, 4.19 could potentially be interpreted as true since the referent of the subject exists, and only the referent of the object (in an opaque position) does not exist; or that, if Sherlock Holmes is taken to be a non-existent entity, 4.19 is truth-valueless since the subject NP refers to a non-existent entity. In this case, both of the possible predictions of the opaqueness explanation, i.e. that the sentence is true or that the sentence is truth-valueless,

---

[10]Other opaque predicates such as **admire**, **worship**, **forget**, and **want** pattern with **dream about** with respect to Sherlock Holmes' brother and Chelsea Clinton's sister.

conflict with the truth-value assigned using native speaker intuition, i.e. that this sentence must be false.

If we consider sentence 4.20 we see that it shares an intuitive truth-value (false) with 4.19 but that 4.21, in contrast, patterns with 4.18.

4.20. Sherlock Holmes dreamed about John Major.

4.21. Sherlock Holmes dreamed about Oliver Cromwell.

To summarise, the opaqueness explanation predicts that only the existential status of the referent of the subject NP of an opaque predicate has any effect on the truth-value of a sentence, and that the location of the existence of the referent of the object NP cannot be important, since the object NP contributes only its intensional meaning. However, we observe that the pattern of truth-values which occurs in sentences 4.15 - 4.21 cannot be accounted for by the opaqueness explanation, and we will argue instead that the object NP *does* refer and further, that the location of the referents of subject and object NPs which occur with opaque predicates influences truth-value assignment[11].

We conclude that there is initial evidence that there are phenomena which cannot be handled by presuppositional theories employing the notion of opaque contexts. From the brief survey of sentences 4.15 - 4.21 it appears that the pattern of truth-values may be bound up with the existential location of the subject and object NPs of these opaque predicates.

4.4.2. Location and Reference Failure.

Just as it is not possible to reason from particular truth-conditions to particular truth-values, neither is it possible to reason from the truth-conditional inevitability of a truth-value in every possible world to the actuality of a truth-value with respect to some particular world. To put

[11]This will not interfere with the failure of Leibniz's law in opaque contexts since the de re/de dicto distinction is retained. cf. chapter 7 for a discussion of how to account for the failure of Leibniz's Law if all referring expressions refer extensionally.

this another way, our knowledge[12] that for any individual it is either true or false that that individual has had a tummy tuck does not automatically enable us to decide whether or not sentence 4.22 represents a true statement.

4.22. Chelsea Clinton's sister had a tummy tuck.

We can say that our understanding of truth-conditional meaning rests on, among other things, our acceptance of the law of the excluded middle (~ (A & ~ A)). Language users, when they infer from the truth of some sentence the falsity of its negation, do so on the implicit assumption that for any particular sentence there can be no possible world that neither meets its truth-conditions nor fails to meet them, and no possible world which both meets its truth-conditions and fails to meet them. When we say that the truth of the sentence in 4.23 entails the falsity of the sentence in 4.24, what we are actually claiming is that the truth of any token of the sentence in 4.23 entails the falsity of any token of 4.24 if and only if the referring expressions in both pick out the same entity.

4.23. The dog decomposed.

4.24. The dog didn't decompose.

Likewise when we say that 4.25 is necessarily false, what we mean is that some token of the sentence in 4.25 is necessarily false if and only if **the dog** and **it** are coreferential.

4.25. The dog decomposed and it didn't decompose.

Despite this, there are many instances of tokens of sentences to which we do not and cannot assign truth-values. To explain this phenomenon, while retaining the notion of a set of possible worlds which adheres to the law of the excluded middle, we could build into our theory (of the application of possible worlds in truth-value assignment) some principle which defines circumstances under which we do not choose *any* possible world as the location of the

---

[12]Whether this is indeed knowledge or should be defined instead as a belief (i.e. whether the law of the excluded middle defines the limits of actual possibility or merely those of epistemic possibility) is discussed in chapter 5.

referent of some expression. This kind of approach to the problem of perceived truth-valuelessness has been taken in the past and still has its followers. It has been suggested by McCawley for example, that while most sentences are "about" particular worlds some sentences should be analysed as being "about" a set of worlds (McCawley, 1981). For example, sentence 4.26 can be interpreted not as being a statement about an entity in some particular possible world, but rather as a statement about the entire set of worlds in which the King of France exists.

4.26. The King of France is bald.

The alleged truth-valuelessness of sentence 4.26 can then be explained in the following way. The proposition in 4.26 reflects a state-of affairs which holds in one sub-set of the worlds 4.26 is about; i.e. those worlds in which the King of France is bald. However, 4.26 does not accurately reflect the complement of this sub-set, the worlds in which the King of France is not bald. Therefore, 4.26 is neither true nor false of the set of worlds in which the King of France exists.

An alternative, although similar analysis (van Fraassen, 1966) discussed by Fodor (1979) takes the view that sentences may be about unidentifiable worlds. In this case, although this sentence must be about one particular member of the set of possible worlds in which the King of France exists, rather than being about the set as a whole, nothing in 4.26 specifies *which* world is the one in question, and therefore we do not know whether it is one of the worlds in which the King of France is bald or not.

A third option which we mention briefly here, and discuss in detail in Chapter 5, is that sentences can be about underspecified worlds. Fodor (1979) proposes an analysis in which a sentence such as 4.26 is about one particular world which is underspecified to such an extent that it only has one property; the existence in it of the King of France. In this underspecified world, then, it is actually neither true nor false that the King of France is bald.

All of these explanations of the truth-valuelessness of 4.26 depend on some kind of indeterminacy with respect to the reference world; either 4.26 is about more than one world, or is about an unidentified world, or is about a world which is itself indeterminate. Since it is sentences like this, indeed this very one, which have traditionally been examined in the presuppositional literature almost to the exclusion of all others, it has come to be accepted that plausible explanations for their truth-valuelessness are plausible explanations for truth-valuelessness in general.

It might be expected that the current theory, which holds that the location of referents is closely (if indirectly) bound up with truth-value assignment, would be well-served by an analysis of truth-valuelessness which sees it as the result of the unlocatability of referents . However, we observe that if we employ this analysis many intuitively truth-valueless sentences cannot be accounted for.

Consider the following sentence:

4.27. Sherlock Holmes has his cheese sent from Sheffield.

Let us assume that 4.27, like 4.26, is perceived as truth-valueless. Given this assumption, we must take one of the two following views. Either the respective truth-valuelessnesses of 4.26 and 4.27 are of different kinds, and therefore may be accounted for by different explanations, or they are of the same kind and are to be explained in the same way. The first of these two options is not attractive. We are wary of positing a plurality of types of truth-valuelessness as anything but a last resort. Let us assume then, that the truth-valuelessnesses of 4.26 and 4.27 are of the same type and let us try to account for 4.26 and 4.27 by applying the three previously presented types of indeterminacy explanation.

An indeterminacy explanation would hold that we cannot assign a truth-value to sentence 4.27 for one of the following reasons:

a. because it is both true and false since it is about a set of worlds, in some of which the proposition that Sherlock Holmes had his cheese sent from Sheffield is true and in some of which the proposition is false.

b. because it is either true or false since it is about some unidentified member of the set of worlds in which Sherlock Holmes exists.

c. because it is neither true nor false since it is about a world which is not specified with respect to the relevant properties.

Explanation a. and explanation b. both predict the correct truth-value assignment for this sentence; we could say that we do not know whether or not it is true that Sherlock Holmes has his cheese sent from Sheffield, because the sentence is not about an individual in any specific world, or because we do not know which specific world the referent of **Sherlock Holmes** inhabits. It is interesting to observe, however, that native speakers seem much more squeamish about employing either of these explanations to account for 4.27 than they do with respect to sentence 4.26. Since neither of these two explanations have any intuitive attraction to add to their empirical adequacy, we will consider explanation c. which states that although **Sherlock Holmes** can be understood as picking out an individual in some identifiable world, i.e. the world invoked by the fiction of Conan Doyle, this established familiar fictional world is incompletely specified. This amounts to a claim that it really is neither true nor false of this world that Sherlock Holmes has his cheese sent from Sheffield. Rather than deciding on the answer to this question on the basis of 4.27[13], let us consider sentence 4.28.

4.28. The Princess of Wales' first nanny had green eyes.

Let us assume that, like the previous two examples, 4.28 is perceived as truth-valueless by current readers. We are equally (if not more) squeamish about accounting for its truth-valuelessness by claiming that it is about a set of worlds, or about an individual in some unidentified member of a set of worlds. Our strong preference is to interpret 4.28 as a

[13]we will deal specifically with the possibility that fictional worlds are incomplete in chapter 5, section 5.3.1.

statement about an individual who exists in the real world. However, we are resistant to the option of accounting for the perceived truth-valuelessness of 4.28 by claiming that the reference world, in this case the real world, is underspecified or incomplete. Not only does this conflict with our intuitions about the nature of the real world, but it also seems unnecessary and irrelevant to complicate the matter in this way to account for the perceived effect.

Instead we can argue that 4.28 is about a particular world and that that world is completely specified with the result that the sentence in 4.28 has some truth-value; but that the inability of hearers (or readers) of 4.28 to assign the appropriate (i.e. the correct) truth-value to it is a result of the particular limitations on the knowledge states of those hearers. In other words, the inability of hearers of 4.28 to assign a truth-value to a statement that the Princess of Wales' first nanny had green eyes arises straightforwardly from their not knowing whether or not the Princess of Wales' first nanny had green eyes. A language user's knowledge that this sentence must be either true or false does not enable them to assign either value to it.

We have been able to account for the truth-valuelessness of sentence 4.28, without complicating the notion of locating referents in possible worlds in any of the ways suggested in other analyses. We can account for the truth-valuelessness of sentences whose referring expressions have potential referents in the real world without being forced to argue that the sentences are not about these real-world entities, or that we do not know whether the sentences are about these real-world entities or that the real world is incomplete. Given this, we see no reason to propose any of the more complicated explanations to account for the truth-valuelessness of 4.26 and 4.27. We propose instead that sentence 4.27 is a statement about an entity in the fully specified world invoked by the fiction of Arthur Conan Doyle, that 4.27 has a truth-value in this world, but that hearers might not be able to assign a truth-value to 4.27 because they do not have sufficient knowledge about the properties of that world. Similarly, we propose that sentence 4.26 is about an entity in a particular fully-specified world and that it therefore has a truth-value. We claim that a statement that the King of France is bald is perceived as truth-valueless because we do not know, and cannot find out, whether or not the King of France is bald.

4.5. Conclusion.

We have given some indication that the relation between the location of referents and truth-value assignment might not be what it has generally been assumed to be. In particular, contrary to the claims of opaqueness analyses, the location of referents of expressions in "intensional" contexts may be crucially involved in the assignment of truth-values, and contrary to the claims of indeterminacy analyses, the location of the referents of expressions in other truth-valueless sentences might have no direct bearing on their truth-valuelessness.

To summarise, for any sentence, there must some world in which the referent of a constituent referring expression can be located such that the sentence is true, and there must be some world such that if the referent is located there the sentence is false. Furthermore, there is no world in which, if the referent is located there, the sentence is neither-true nor-false or both-true-and-false. However, tokens of these sentences undergo only one truth-value assignment and that might result in the assignment of any value. A theory which predicts whether a particular utterance of a sentence is actually true, actually false or neither, must resolve the question of how we decide in which of the infinitely many possible worlds where we could locate the referent of some expression, we actually do so. However, although the correct choice of reference world is necessary, it alone may not be sufficient to account for truth-valuelessness. We will analyse truth-valuelessness as the result the failure of a part of the interpretation procedure which is separate from, and subsequent to, the fixing of referents in some world.

This procedure, which we set out in the chapter 6 of the thesis, depends, as stated earlier, on a particular view of the nature of possible worlds which allows the imposition of a particular structure on the set. In the next chapter we discuss the nature and ontological status of possible worlds and set out the structural relationship that holds between them.

Chapter Five.
The Set of Possible Worlds: Arrangement and Relations

5.1. The Current World.

We proposed in the previous chapter to define one world and then relate all other members of the set of possible worlds to it. This world, the most possible world, will henceforth be labelled w_0. For the purposes of the following discussion w_0 is assumed to be the world in which this thesis is being written, unless it is explicitly stated otherwise. We will argue that in general the current w_0 should be defined relative to individual acts of utterance or discourses such that if an utterance is made in or with respect to a particular world then that world is w_0 for that utterance. Thinking of w_0 as the current world rather than the real world allows us to conceive more easily of shifts in its identity. To suggest that the identity of the real world changes when we consider different discourses could possibly lead to confusion since there is a strong intuitive pull to the notion that the identity of the real world is fixed[1]. On the other hand, we want to be able to interpret an act of utterance which takes place in another world on the assumption that that other world is w_0 for that utterance. This is essential if we are to understand deixis in fiction and in counterfactuals. For example, when we read that Scarlett O'Hara said "*I can't think about that today; I'll think about that tomorrow*", we understand that the proposition expressed is that Scarlett O'Hara will not think about some issue which she previously mentioned on the day she makes this utterance but will instead think about that issue on the subsequent day. We do not infer that (to take one possible example) Margaret Mitchell decided not to think about something that the typesetter of the edition of the novel that we are reading said over lunch on the day that we read the sentence[2].

---

[1]In the novel Sophie's World by Jostein Gaarder the heroine discovers that the characters she is reading about in a book are real and that it is she, herself, who is fictional. The implication is that we, the readers of Gaarder's book, might make the same discovery.

[2]Indeed, our acceptance that deixis works in fiction according to a locally determined w_0 is so strong that it is questionable to label the second interpretation "possible" at all.

5.1.1 The Set of Possible Counterparts to the Current World

Given the current world, a modal realistic approach demands that every possible counterpart of that world exists. That is, every world which differs from w_0, by virtue of any or all of its properties (items and attributes) co-exists with w_0 in the set of possible worlds. Since the theory of reference and truth-value assignment which we are working towards in this thesis, involves the choice of one over all other possible worlds as a location for the referent of linguistic expressions, and since modal realism stipulates that all worlds are of the same type, to make any sense of the process of choosing a world we must impose a structure on the set of worlds. That is, since all worlds are of equivalent status and cannot therefore be distinguished by virtue of their inherent type, we must distinguish between worlds by means of the relations which hold between them. We claim that the set of possible worlds is such that given some w_0 all other worlds can be ordered with respect to their proximity to it. Following Lewis (1968, 1986) we will assume that non-real "versions" of the real world are best seen as counterparts of that world, and that the properties of these counterpart worlds, the sets of entities and relations, are counterparts of the properties of the real world. We do not claim that *the same* entity may exist (or *the same* relation hold) in two separate worlds (McCawley, 1981; Lewis, 1986)[3].

The first feature of the set of worlds we propose is the relation of proximity between counterparts. The proximity of some world w_0' to another world w_0 is measured by the number of counterpart properties which have a different value in w_0' and w_0. We can say that given the current world w_0 and some other world w_0''' there is some number of intervening worlds, which can be ordered with respect to decreasing proximity to w_0''' and increasing proximity to w_0. We can say that w_0 and w_0''' are connected by a string of intervening counterparts with as many intermediates along it as there are differences in the values of the counterpart properties of w_0 and w_0'''. For example, if there is a world w_0 such that in w_0 some individual *a* has the properties of being single, childless, and a student then

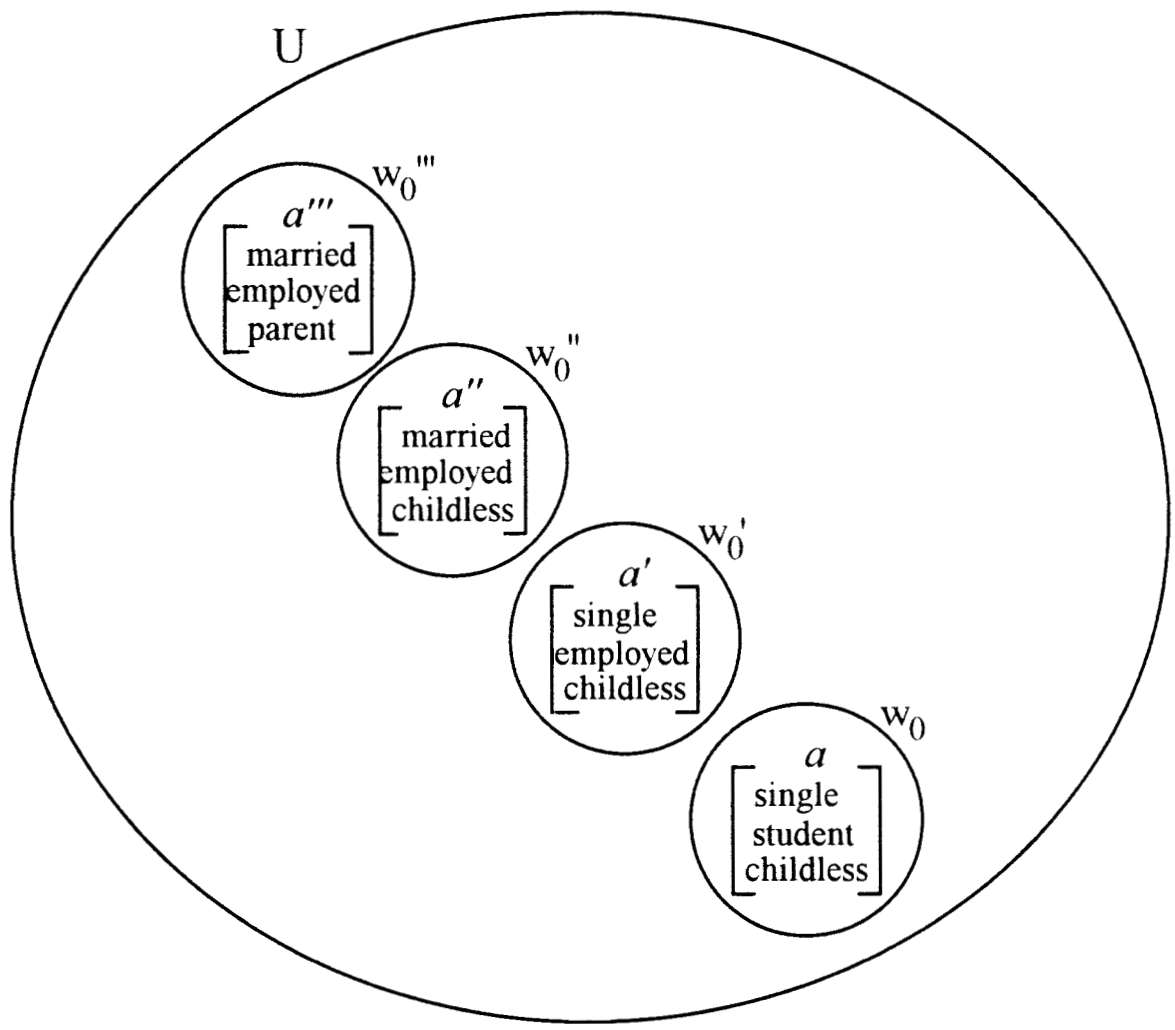
---

[3]There are concomitant difficulties with the notion of counterparts, especially with respect to the use of names in modal statements c.f. Kripke (1980), but since our interest is primarily in the use of linguistic expressions we are content to reflect what appear to be general intuitions about the use of names (chapter 6 section 6.2 and chapter 7 section 7.2.2) not withstanding possible philosophical objections.

there is a world w_0' in which a' is employed, single and childless, a world w_0'' in which a'' is a employed, married and childless and a world w_0''' in which a''' is an employed, married parent. Figure 5.1 represents a partial diagram of the set, U, of possible worlds showing this string of counterparts.

Figure 5.1

A partial diagram of counterpart worlds

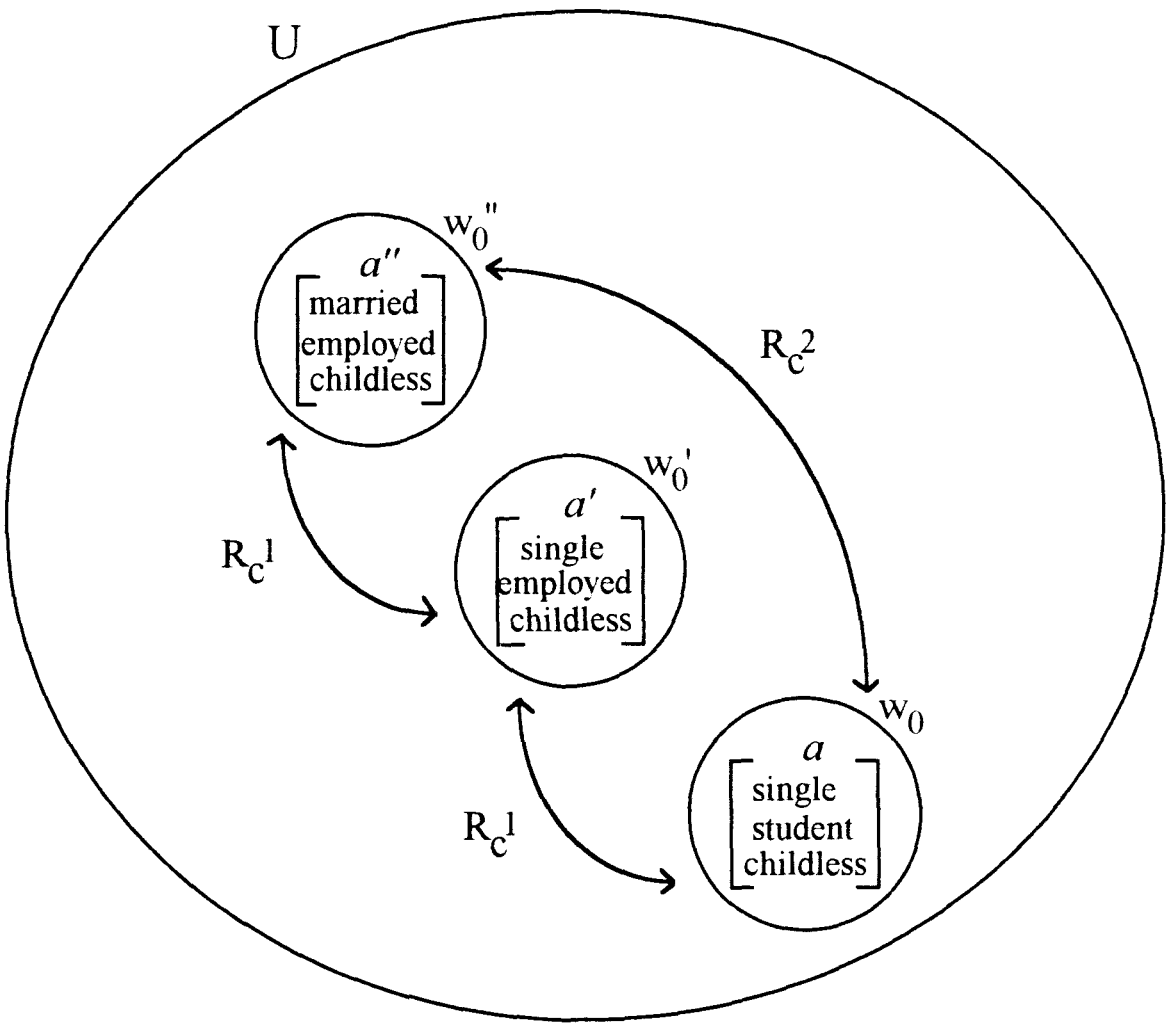


From now on we will describe strings of worlds in terms of a counterpart relation, R_c. A one-place counterpart relation R_c^1 is a relation which holds between some world w_0 and some world w_0' which is an identical counterpart of w_0 except for the value of one property or minimal bundle of properties. The use of minimal bundles of properties, rather than single properties allows us to define w_0' in fig 5.1. as a one-place counterpart of w_0 even though , if the counterpart of a in w_0' has the property of being married which the counterpart of a in w_0 does not, then some other individual b in w_0' has the property of being married to a', some third individual has the property of having married a' to b and so on. A two place counterpart link is a relation R_c^2 which holds indirectly between two worlds w_0 and w_0''

which differ from each other by the value of two minimal bundles of properties, and which are each joined by a one-place link R_c^1 to an intervening world w_0'. Thus, we can define R_c as a cumulative relation such that if $w_0\ R_c^1\ w_0'$ and $w_0'\ R_c^n\ w_0''$ then $w_0\ R_c^{n+1}\ w_0''$. We can also define R_c^n as a symmetric relation such that if $w_0\ R_c^n\ w_0'$ then $w_0'\ R_c^n\ w_0$. Figure 5.2. shows the relations between a partial string of counterparts in U.

Figure 5.2

A partial diagram of counterpart worlds showing counterpart links



This relation gives us a way to express the notion that every world is related to every other world. We will propose further inter-world relations in the forthcoming discussion. Henceforth we will call a world w_0' which is linked with some w_0 *only* by an n-place counterpart relation, a counterpart of w_0. We will call a world w_0' which is linked to a world w_0 by a one-place counterpart relation a minimal counterpart of w_0.

5.2. Temporal Counterparts of the Current World.

It is relatively straightforward to characterise the elements which co-exist spatially in w_0 in a comprehensible way; we have no difficulty in understanding definitions such as Bradley and Swartz (1979), McCawley (1981) or Lewis (1986) that w_0 consists of everything that there is, or the totality of actually existing states of affairs, or the universe as a whole etc. Much more problematic, however, is the question of what, if any elements, co-exist temporally in w_0. That is, whether w_0 is an instantaneous snap-shot of all at that there is at some moment or whether w_0 contains a temporal dimension. Philosophers, logicians and linguists disagree fundamentally on this point. For example, Bradley and Swartz define the real world as "all that was, is or will be" (Bradley and Swartz, 1979: 9), firmly indicating that they take worlds to have a temporal dimension, while the proponents of tense logic assume that moments in time are the temporal equivalents of the wholly spatial worlds of modal logic (McCawley, 1981). Lewis argues that worlds persist through time in a way which he calls "perdurance" which is the temporal equivalent of the way that entities such as roads persist through space; different parts occurring/existing at different temporal/spatial locations (Lewis, 1986) .

Since we proposed in chapter 4 that two such entities as John Major and Pitt the Younger exist in different worlds, and since their respective worlds can be seen as distinct from each other only temporally, it appears that we are assuming the stance that worlds do not contain a temporal dimension. We must now look closely at the question of whether w_0 should identify the current world as it is now or the current world at all times. Much of the difficulty in this area concerns the problem of the future. There is a great deal of debate over whether the future is the same kind of thing as the past. It is not a clear-cut issue, in fact, whether there is *a* future at all, and not a set of indistinguishable potential futures. Since these questions are of a kind which do not generally arise when we consider the past, we will take the expedient step of discussing earlier and later versions of the current w_0 separately.

5.2.1. The past.

If worlds include no temporal dimension, but are instantaneous snapshots, then we must add to their ontology a temporal index, such that w_0 consists of a set, A, of entities, a set, P, of attributes and a time index, t. We cannot take the temporal index of a world to be one of the properties which make up the set of its attributes since this would leave us unable to distinguish those worlds which are "earlier versions" of w_0 from other "alternative versions" to w_0 which differ from w_0 in other ways. That is, each of the three worlds, about which 5.01 - 5.03 can be read as true statements, would stand in the same relation to the other two.

5.01. The Queen has advised an early divorce.

5.02. Charles and Diana have announced their engagement.

5.03. Diana has broken off the engagement after a chat with Camilla.

One way to avoid this is to say that 5.01 and 5.02 are about the real world at different times and that 5.03 is about some other world. However, if the real world as it is in 1996 is taken to be a different world, w_0, from the real world as it was in 1796, say $w_{0\text{-}200}$[4], we must impose some structure on the set of worlds, which will identify these two worlds, and all those worlds which intervene, as members of an internally consistent sub-set.

Ideally we would like to be able to define earlier versions of w_0 as those possible worlds which are identical to the real world save that they have a lower time index. If $w_0 = \langle A, P, t_0 \rangle$, then we could define an earlier version of w_0, e.g. $w_{0\text{-}200}$, as $\langle A, P, t_{-200} \rangle$ and any earlier version of w0, w_{-n} as $\langle A, P, t_{-n} \rangle$. However, there is clearly a problem with this definition, in that many attributes of a world bring about change in the inventory of items and/or attributes in that world. To put this another way, a true statement about some entity in a world might entail that the membership of the sets of items or attributes in that world

[4]The use of the subscript -200 to distinguish 1996 from 1796 is for ease of interpretation only. We do not mean to imply that only 199 temporal counterparts intervene. Since the problem of identifying the smallest interval of time is currently unsolved, the value of this index is arbitrary.

changes. Such attributes include the processes expressed by change-of-state verbs such as **die**. For example, let us assume that sentence 5.04 represents a true statement about an entity in some world.

5.04. Pitt the Younger died.

Our intuitions tempt us to define this world as w_{-190} with respect to w_0 in the current context. That is, we take sentence 5.04 to be a statement about an individual in an earlier version of the current world. However, the individual named, Pitt the Younger, even if he exists as a member of the set of items in w_{-190}, does not exist in w_0, therefore by the definition above w_{-190} is not an earlier version of w_0. This is not an insuperable problem, however. We must abandon the attempt to define temporal counterparts as worlds which are identical counterparts apart from their temporal indices and instead define them as worlds which are temporally linked and mutually consistent with one another. We can then refine the definition of "an earlier version" as follows: w_{-1} is an immediate (one-place) temporal counterpart of w_0 iff w_{-1} differs from w_0 only by having a lower time index and by such properties as are consistent with the entities which exist and the attributes which hold in w_0 and w_{-1}. Thus, w_{-n} is an n-place temporal counterpart of w_0 iff w_{-n} differs from w_0 only by having a lower time index and by such properties as are consistent with the entities which exist and the attributes which hold in w_0, w_{-n} or any intervening temporal counterpart.

Clearly the value of this definition is bound up with the possibility of defining consistency. The definition of consistency for a logical system is straightforward. For example:

> "A formal system is consistent if it is not possible to derive from the axioms of that system both some statement and the denial of that same statement."
>
> (Partee, ter Meulen and Wall, 1990: 202)

If, therefore, we were to view worlds as linguistic or logical constructs, e.g. if a world was seen as a set of propositions, it would be a simple matter to distinguish pairs of consistent worlds from pairs of inconsistent worlds. However, since we take a modal realistic view of worlds, the problem is more serious. Can we say that a pair of worlds is temporally consistent

iff it is not possible to produce a contradiction by combining true statements about each world? Clearly we can state of 1796 that Pitt the Younger exists, and state of 1996 that Pitt the Younger does not exist. The only way to show that a conjunction of these two statements is not a contradiction is to appeal to the intervening counterparts. This, however, would render our definition circular.

We propose to break the circularity involved in stating a formal definition of temporal consistency and inconsistency for pairs of worlds, by the use of paradigm examples. Below we cite what we take to be a clear case of a set of temporally consistent worlds and a clear case of a set of temporally inconsistent worlds. We acknowledge that there will be many less clear examples. However, we can observe that those worlds which appear to be borderline cases, as regards their temporal consistency with the current world w_0, are worlds either in the future[5] of w_0 or in the very remote past. For example, there is more general agreement about the identity (the properties) of 1995, than there is about the identity of 1997 or of 20,000 BC.

Recall that the modal realistic view of possible worlds taken in the current thesis applies only to the worlds themselves. We do not claim that the structure on the set is anything other than a theoretical construct developed to allow us to model interpretation. Given that our view of interpretation is based on the processes which take place in real discourse contexts involving speakers and hearers, there is no difficulty in accepting that the limits to speakers' and hearers' knowledge about worlds affects their ability to interpret statements about these worlds. In short, if we accept that the borderline cases lie in the future and the remote past, then it is surely a positive feature of an epistemically based theory that these worlds *are* borderline cases.

Let us take the following three worlds as a clear example of temporal consistency:

$$w_{-200} = \langle A, P, t_j \rangle$$
$$w_{-190} = \langle A, P', t_k \rangle$$
$$w_0 = \langle A', P, t_l \rangle$$

---

[5]we return to the topic of limit epistemic access to the future in section 5.2.2.

where A and A' (the sets of items), are identical except that Pitt the Younger is a member of A but not a member of A' and P and P' (the sets of attributes) are identical except that Pitt the Younger has the property of dying according to P' but not according to P. In other words, in w_{-200} Pitt the Younger exists, in w_{-190} Pitt the Younger dies and in w_0 Pitt the Younger does not exist. We assume for the sake of simplicity that all other properties of the three worlds are identical.

w_{-200} can be seen to be a temporal counterpart of w_0 since it differs from w_0 only by features which do not block a relationship of succession between worlds: w_{-200} has a lower time index than w_0 and A in w_{-200} has a member which is not a member of A' in w_0 *but* the non-existence of that item in w_0 is entailed by an attribute of the intervening world w_{-190}. w_{-190} can in turn be identified as an temporal counterpart of w_0 since it differs only in terms of having a lower index and such differences in the membership of their respective sets of items and attributes as are consistent with properties of w_0 (since every property must be consistent with itself). Similarly, w_{-200} can be identified as a temporal counterpart of w_{-190}.

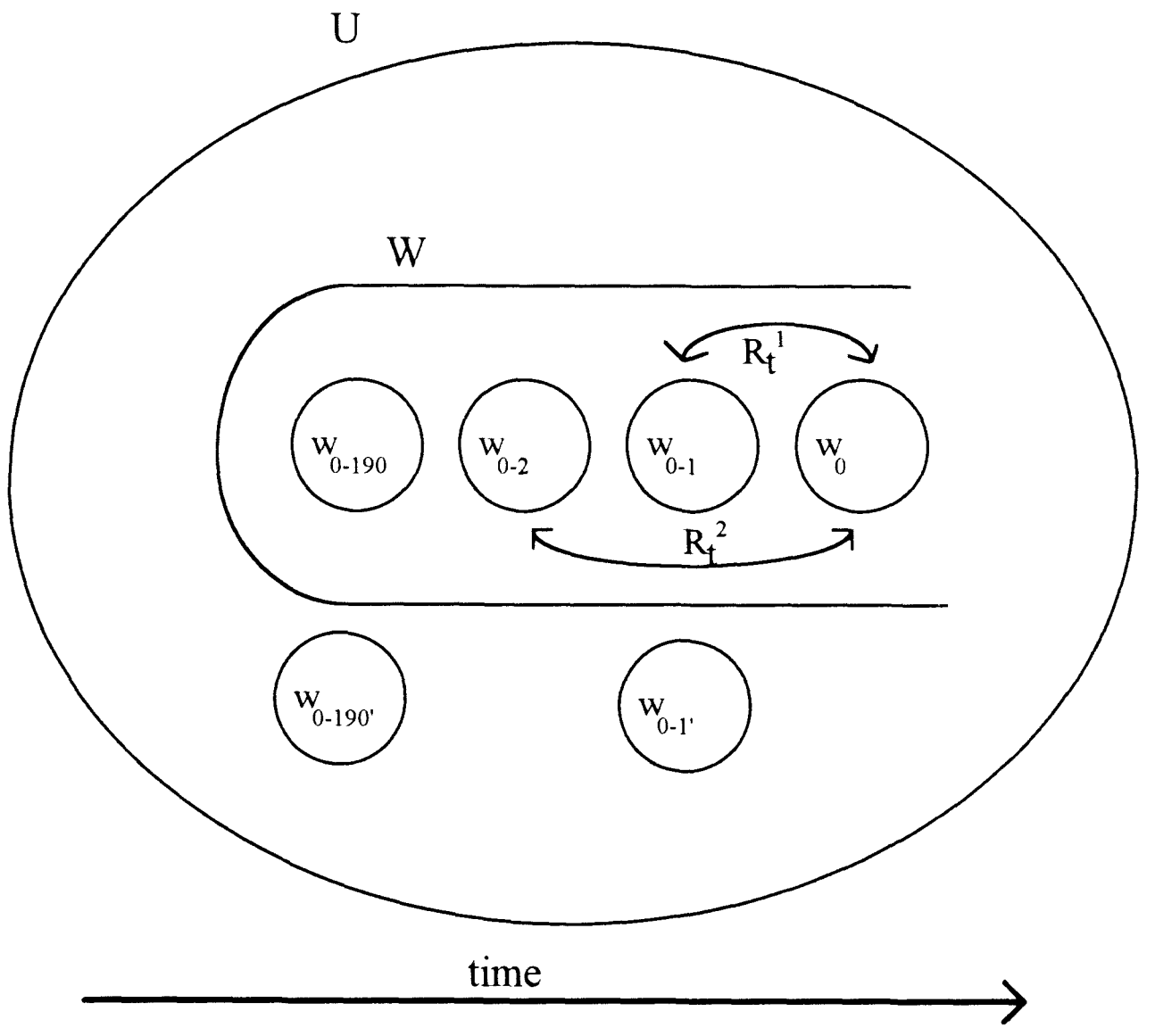
In contrast, let us consider a similar set of three worlds: w_{-200}, w_{-190}', and w_0. w_{-190}' is identical to w_{-190} except that in w_{-190}' Santa Claus is a member of A. Since no attribute of w_{-200} brings about the existence of Santa Claus, and there is no intervening world between w_{-200} and w_{-190}' with such an attribute, w_{-190}' is not consistent with w_{-200} and cannot be one of its temporal counterparts. Similarly, since no attribute of w_0 (or any other world which intervenes between w_{-190}' and w_0) brings about the loss of Santa Claus from the membership of either of the sets of items A or A', w_{-190}' is not consistent with w_0 and cannot be one of its temporal counterparts.

We will call the set containing w_0 and its temporal counterparts W. Given some w, we can identify a partial membership for the corresponding W, using a relation R_t which defines what we think of as subsequent, temporal ordering or chronological succession. A one-place temporal link R_t^1 is a relation which holds between some world w_0 and w_{0-1}; i.e. between a world and an immediately preceding temporal counterpart of that world. An n-place temporal

link R_t^n therefore, holds between a world w_0 and a world w_{0-n}. R_t cannot be described as a transitive relation, since if $w_{-2}\ R_t^1 w_{-1}$ and $w_{-1}\ R_t^1\ w_0$, then it is not the case that $w_{-2}\ R_t^1\ w_0$. We can describe R_t instead as a cumulative relation since if $w_{-2}\ R_t^1 w_{-1}$ and $w_{-1}\ R_t^1\ w_0$, then $w_{-2}\ R_t^2\ w_0$. Figure 5.3. shows a partial diagram of the set, U, of possible worlds, containing part of the set W associated with the current w_0.

Figure 5.3

A partial diagram of W for the current w_0



To summarise: $w_i\ R^t_1\ w_j$ iff $w_i = \langle A, P, t_m\rangle$ and $w_j = \langle A', P', t_n\rangle$ such that $t_m = t_{n-1}$, A is consistent with A' (with respect to P'), and P is consistent with P' (with respect to A')[6].

---

[6]This view of the association of time and worlds departs from the formulation in Montague Semantics (Dowty, Wall and Peters, 1981; Cann, 1993) in that worlds for Montague are not

We can differentiate the worlds in W either by naming them, as in figure 5.3, using calendar dates or by identifying attributes which distinguish them from all the other depicted worlds (although not from all other possible worlds). w_0, 1996, is the world in which the planetary probe Discovery enters Jupiter's atmosphere. w_{-1}, 1995, is the world in which the Queen advises an early divorce. w_{-2}, 1994, is the world in which the Soccer World Cup takes place in the USA. w_{-190} is the world in which Pitt the Younger dies. The other worlds in U cannot be named using calendar dates but can be identified by means of their attributes. $w_{0\text{-}1}'$ is the world in which the Queen enters Jupiter's atmosphere, and $w_{0\text{-}190}'$ is the world in which Pitt the Younger discovers the structure of DNA.

Because the definition we have given of temporal counterparts states that a temporal counterpart of some world w_0 must have a *lower* time index than w_0, R_t is an asymmetric relation which can only be used to relate some w_0 to earlier versions of itself but not to later versions of itself. We must now decide whether we can define later versions of w_0 in the same way which would allow us to re-formulate R_t as a symmetric relation and close the open-ended set W.

5.2.2 The Future

As was mentioned briefly in the previous section, there is a great deal of debate about how to account for statements about the future. Lyons (1977) discusses the view that the future is fundamentally different from the past and argues that this difference can be summed up by the claim that "futurity is modal in nature".

> "Ever since Aristotle first raised the question, the factuality of statements descriptive, or predictive, of future events, or states-of-affairs, has been philosophically controversial and many philosophers would deny that we can make statements about the future at all, on the grounds that we cannot have knowledge, but only beliefs, about future world-states."
>
> (Lyons, 1977: p814-815).

---

temporally bound. Instead, W is constructed as a set of indices <w,t> which are world/time pairs. W is ordered with respect to the set of all temporal indices T, using a precedence relation <, such that $\forall t \in T, \exists w\ [<w,t> \in W]$. So what is an index in MG, we are taking to be a world.

One side of the argument then can be summed up as follows: a factual statement is a statement which can be assigned a truth-value, and truth-value assignment requires verifiability. Since we have no access to the future, attempted factual statements about the future cannot be verified and so are closer in status to modal statements, which are also not directly verifiable, than to successful factual statements about the present and the past. This point of view crucially depends on the claim that there is more than one possible future and the actual future is not identifiable.

The opposing stance is taken by Lewis, who observes that since we can make statements about the future there must be only one future since this alone allows us to explain what it means to talk about it. (Lewis 1986: 207). We assume that what Lewis means by making a statement is what Lyons means by making a factual statement, i.e. making a statement with a verifiable truth-value.

The disagreement can be summarised as whether there is one future which is the same kind of thing as the past, or a branching future which can be contrasted with a unified past. Lewis is strongly opposed to any analysis which might force the acceptance of the possibility of overlap between worlds. He therefore argues against the notion of a branching future since, he claims that this would result in the possibility that two worlds overlap (Lewis, 1986). He argues that given the idea of a branching future,

> "...one world consists of the initial segment plus one of its futures; the other world consists of the identical initial segment plus the other future. In divergence on the other hand there is no overlap. Two worlds have two duplicate initial segments, not one which they share in common."
>
> (Lewis 1986: 206).

The difficulty we perceive with the notion of the divergence of two worlds with duplicate initial segments as an explanation of the future, is that it seems to imply that what happens in the future defines the present. If we take Lewis's future to be w_0 then what he seems to be saying is that the identity of the w we currently inhabit depends on what happens in the future. If, for example, our alarm clock does not go off tomorrow morning then we currently inhabit some w which is a temporal counterpart of that w_0. On the other hand, if our alarm clock

does go off tomorrow morning, then we currently inhabit a *different* w, which is a temporal counterpart of *that* w_0. Given Lewis's view that overlap between worlds is impossible, this is not a trivial claim. It means that the w we currently inhabit really is a different entity depending on what will happen in the future. Lewis appears to be saying that the future is determinate but the present is not. This is an extremely counter-intuitive notion.

A solution to this problem is possible given the distinction we have proposed between w and W. We can retain the notion of the indeterminacy of the future without having to countenance the possibility of overlap, in either of two ways. We can assume that we cannot identify the complete membership of the W we inhabit, since it contains counterpart ws which correspond to alternative future states of affairs. This is the view of W which is represented in the diagram in Figure 5.4. Alternatively, we can say that each possible W contains only one consistent set of temporal counterparts but that we do not know which W we inhabit. This view of W is represented by the diagram in Figure 5.5.

Figure 5.4 A single W containing counterpart futures

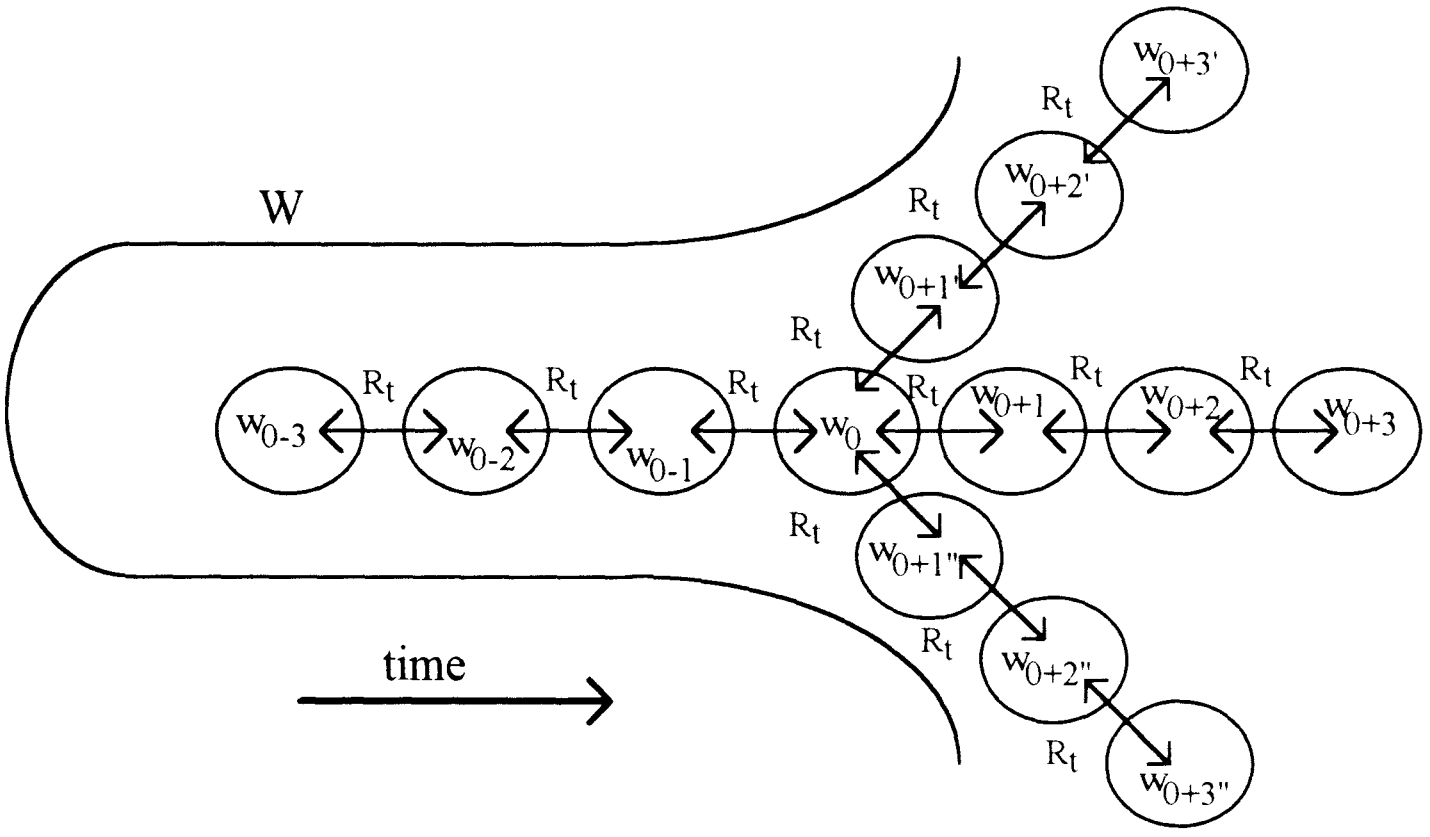
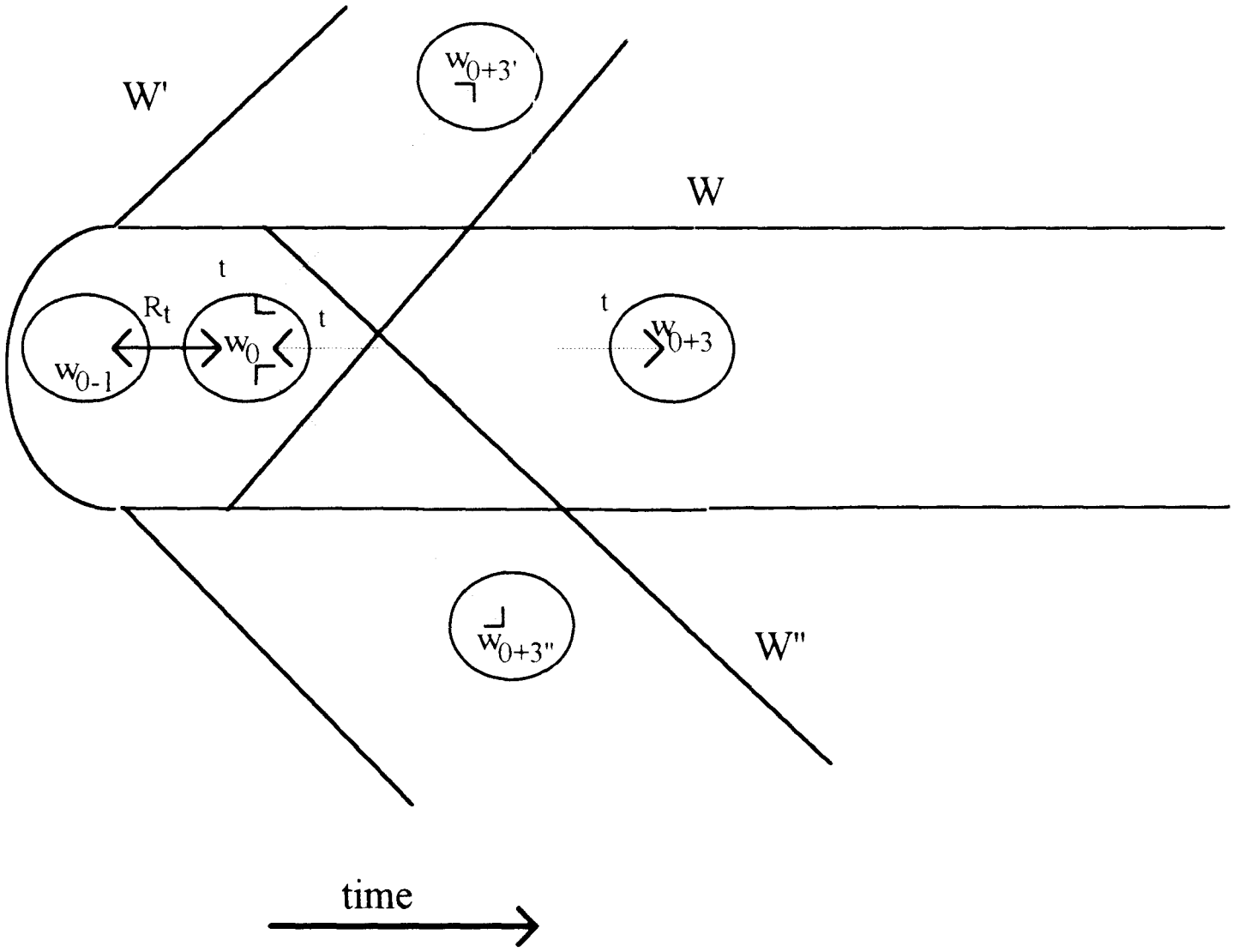


Figure 5.5 A partial diagram of three internally consistent W's



Lewis goes on to argue that "a modal realist who believed in genuine branching in which his world overlaps with others by having initial segments in common would agree (that the future has no determinate truth-value - CMcP)." (Lewis, 1986: p207). Given the second of the views of W sketched above, i.e. that in Figure 5.5., understood in a modal realistic sense, we believe that it is possible to hold that the future branches while at the same time, claiming that the future is determinate. We can say that the future branches in that it contains infinitely many possible alternative ws. Since each of these ws, however, exists in a consistent W made up entirely of temporal counterparts, we can say that every W and every w is determinate, therefore the particular W which we inhabit and all of its constituent ws are determinate. This does not necessarily contradict Lewis's claim however, as it might be that this view of the structure of W does not constitute "genuine" branching.

While we deny that our view of W forces us to accept that the future has no determinate truth-value, we happily concede that the future has no determinable truth-value from our vantage point in our present, which is the future's past. Future versions of w_0 can be thought of as further examples of worlds which exist and are completely specified but about which our knowledge is limited. The only relevant difference between the reference worlds we use for 5.05 and 5.06 is that we might soon come to find out whether 5.05 is a true statement about the world in which the referent of the definite exists whereas the reference world employed for 5.06 is more likely to remain inaccessible.

5.05. Tomorrow the sky will turn red.

5.06. The Pope's wife wears lovely hats.

Lewis argues that the fact that we make statements about 'the' future means that the idea of a branching future is flawed. He asks what it would mean to say that "The future holds a sea-fight." if there were more than one future. (p207) We, on the other hand, have no difficulty with the idea that it is possible to make assertions of propositions to which neither the speaker nor any hearer is able to assign a truth-value. A speaker can make the assertion in 5.07

secure in the knowledge that they may be mistaken about what 'the' future is; i.e. which future will turn out to *be* 'the' future.

5.07. The future holds my death at fifty.

Furthermore, given a thirty year old speaker, hearers may not be able to assign a truth-value of *true* to 5.07 for twenty years, although they may be able to assign the truth-value *false* sooner, if the speaker is mistaken about the future and dies before the age of fifty, but we do not see how this fact would prevent anyone from knowing what it means for the speaker to make the assertion. Roughly, 5.07 can be paraphrased as 5.07'.

5.07'. I assert that out of the many possible futures the one which will come to pass is the one in which I die at fifty.

The apparent problem with this and other future statements dissolves if we distinguish the act of making an assertion from the act of assigning a truth-value to it. Lewis appears explicitly to reject this kind of distinction. He claims that,

> "...respect for common sense gives us reason to reject any theory that says that we ourselves are involved in branching . . . But we need not reject the very possibility that a world branches. The unfortunate inhabitants of such a world, if they think of 'the future', as we do, are of course sorely deceived, and their peculiar circumstances do make nonsense of how they ordinarily think. But that is their problem; not ours, as it would be if the worlds generally branched instead of diverging."
>
> (Lewis, 1986: p209).

We claim, in contradiction to Lewis, that the inhabitants of the current w_0 in which this thesis is being written *are* in the unfortunate position he identifies. We make statements about the future, some of which turn out to be true and some of which turn out to be false, but none of which can be assigned any truth-value at all at the time of utterance. We do behave mistakenly as if there is only one possible future when we predict states-of-affairs or events which do not occur, when we feel happiness in the anticipation of pleasant future events which

fail to take place, when we feel dread in the anticipation of unpleasant future events which do not occur.

We have been content to recognise the existence in the set of possible worlds as a whole, worlds to which we have permanently and severely restricted access, such as the reference world employed in interpreting sentences about the King of France. Therefore we see no reason to balk at the inclusion in W of ws to which we currently have no access. That is, we see no reason not to include future temporal counterparts in the membership of W as well as past temporal counterparts. As we indicated above however, there are two possible ways to do this and we must now decide between them.

The first approach takes what we might call the omniscient view of W on which W contains the actual future and the actual past. The second approach, which we might call the epistemically constrained approach would include in W the actual past and all possible futures. On either approach our decision to include future temporal counterparts in W has the result that our knowledge about W is limited in some way, since even though future counterparts are entities of the same type as past counterparts (i.e. even though the future is the same kind of thing as the past) they are viewed by us from a different perspective. The problem is that we cannot achieve the perspective which allows us the same kind of access to the future as we have to the past.

If W contains only the actual future and the actual past we have to accept that we do not know which of a number of potential W's we inhabit. If W contains a branching future then we can say that we do know which W we inhabit but we do not know what the properties of W are. The second of these two options would more closely resemble the argument we put forward regarding our of use of counterparts as reference worlds for expressions which occur in sentences we perceive as truth-valueless. There we claimed that we can locate the correct world but that we do not have sufficient knowledge about its properties to assign truth-values to all statements about entities which exist in it. However, the problem with arguing that we can in a similar way, locate the W we inhabit but cannot identify the constituent ws which it contains, is that we cannot then claim that future temporal counterparts are related to w_0 in

the same way as past temporal counterparts. That is, we cannot claim that W is internally consistent if it contains alternative futures. For this reason, we choose to analyse the membership of W as a complete set of temporal counterparts and we must conclude that our lack of access to future ws means that we do not know which W we are in.

5.2.3 Constraints on the Past and Future.

Although we must concede that we do not know which W we are in, we will now consider the question whether we can say what Ws we are *not* in. That is, we will attempt to establish what, if anything, constrains the past and the future; whether we can make a distinction between possible and impossible pasts and futures.

It is deceptively straightforward-seeming to harness our intuitions about what distinguishes possible from impossible pasts. Quite simply, it appears to us that what is possible in the past is what exists there and nothing more; i.e. the only possible past is the actual past. In contrast, with respect to the future, it is what intuition tells us is possible that constrains what we will accept is going to take place or exist; i.e. it appears that the actual future is one of the possible futures. This distinction is, however, spurious as we can see if we consider more carefully the reliability of our intuitions about the future and the past. These are not so clear cut in relation to past and future temporal counterparts which are far removed from w_0. Although our epistemic access to the past is not limited in the striking way that our access to the future is, it *is* limited and our access to more and more remote pasts is increasingly restricted. When we try to access temporal counterparts of w_0 which are among those we describe in terms of geological time, they look surprisingly like the future in that what we take to exist there is constrained by what we consider possible. It is important, therefore, to divorce the question of our constrained access to the future and the past, which appears to allow indeterminacy, from the indisputable fact there is only one set of temporal counterparts in the W we inhabit. The question which remains is whether there are any worlds which cannot be temporal counterparts of this w_0.

We can immediately identify worlds which are inconsistent with earlier members of W as ws which cannot be temporal counterparts of the current w_0 either in the past or in the future[7]. However, we cannot so easily determine what ws are inconsistent with future temporal counterparts of w_0 since we have no access to those temporal counterparts which will allow us to identify their properties and from this to identify what properties are consistent with them.

Although we are less sure about what is possible in increasingly remote futures (and past), it may be possible to set some outer boundaries. We will consider the plausible contention that the set of possible future ws must be contained within the set of physically possible worlds. We must at the outset distinguish two types of worlds which could each be thought of as physically possible. There is a set of worlds whose membership can be defined as those worlds which are consistent with the contingent physical properties of w_0; that is the items and attributes which make up w_0. There is, besides, a larger set of worlds, which contains these and also other worlds which are not consistent with the contingent physical properties which hold of w_0 but adhere to the physical *laws* of W. For example, with respect to the current w_0, a world w_0' which contains the pair of individuals about whom 5.08 represents a true statement, is a physically possible world in both senses.

5.08. Michael Jackson is divorcing Elvis Presley's daughter.

5.08 can be used to make a true statement about a world w_0' which contains only those entities which exist in w_0 of whom only those properties which hold in w_0, hold, since Michael Jackson and Elvis Presley's daughter exist in w_0 and are married to each other. In contrast, again with respect to the current w_0, the world in which Sherlock Holmes exists is a physically possible world in the second sense in that it does not flout any physical laws which pertain in w_0, but is a physically impossible world in the first sense in that its properties are inconsistent with the properties of w_0. The world w_0''' in which Mary Poppins exists is a physically impossible world on both counts, since as well as its inconsistency with w_0, it

[7]always remembering that the more remote the earlier temporal counterpart is the less reliably we can determine what is and what is not inconsistent with it.

flouts several physical laws which pertain in w_0, such as the physical laws which relate an organism's anatomy to its ability to fly while subject to the force of gravity.

We can say that the narrower definition of a physically possible world picks out the set which contains the possible earlier temporal counterparts of w_0 in any W. The question remains whether the broader definition of a physically possible world picks out the set which contains all future temporal counterparts of w_0 for any W. The answer to this question hinges on the reliability of physical laws. If we accept that the particular physical properties in the current w_0 are contingent and we accept that what we call physical laws are observations based on the evidence of the properties which hold in w_0, then we can conclude that the physical laws are reliable to the extent that our knowledge of the properties of w_0 is complete. However, since there are certain facts about w_0 (for example about parts of the universe outside our solar system) that we do not currently know, and that we know that we do not know[8], it is not valid to argue that our epistemic states are reliable indicators of physical possibility. We conclude that the identity of W can not be narrowed down according to what we take to be the physical laws of w_0, although we can be fairly sure that close future temporal counterparts of w0 will probably adhere to current physical laws.

5.3 Possible and Impossible Worlds.

Could we say then that all we know is that W contains a set of successive future temporal counterparts of w_0 and that these are contained in the set of possible worlds. Can we, in fact, even say this much? Is the set of worlds equivalent to the set of logically possible worlds, and if not must W be contained in the smaller set?

---

[8]Consider the discovery (reported in the press in on January 18, 1996) of planets which orbit the stars 70 *Virginus* and 47 *Ursae Majoris* whose moons *may* have atmospheres containing liquid water and other essential features for the development of life similar to that found on Earth. These planets are only the third and fourth known planetary systems (including our own Sun's); we do not currently know if conditions are suitable for life to develop in these planetary systems, and we do not know how many, if any, other planetary systems there are in our own Galaxy.

We can define a logically possible world w_0''' with respect to the current w_0, as a world which has no properties that contradict the laws of logic which obtain in w_0. For example, by this definition, any world which has a pair of contradictory properties such as can be truly described by the statement (A & ~A) is not a logically possible counterpart of w_0 in which the law of the excluded middle stipulates ~ (A & ~A).

As with physical laws, we must decide whether to view the division between logical possibility and logical impossibility as a real distinction (in the modal realistic sense), or as an artefact of our incomplete knowledge. That is, we must decide whether logical possibility is a property of everything that is, or whether there is a set of worlds which have the property of being logically impossible; i.e. whether the set of all logically possible worlds is actually a sub-set of the set of all worlds. This view of possible worlds rests on the notion that logic is a function of human psychology. If it is, then the location of the boundary between what is "logically" possible and what is "logically" impossible is a contingent feature of our particular w_0 and therefore cannot delimit W.

Bradley and Swartz (1979), whose conception of a world is as an abstract construct, nevertheless claim that the boundary between possible and impossible worlds is real[9]. They argue that the limits to possibility are not a function of human psychology since possible worlds are not constrained by (nor are they obliged to extend to) the limits of conceivability and argue vehemently against the opposite view which they label "psychologism" (Bradley and Swartz, 1979). We find their argument unconvincing. To support the claim that logical possibility is distinct from conceivability they appeal to the historical example of ancient mathematicians attempting to square the circle, a procedure which Bradley and Swartz point out "we now know, and can prove, . . . is wholly beyond the bounds of possibility." (ibid. p.3). The problem with this argument lies with the reliance on "proof". "To prove" in this context means to show that something is the case to the satisfaction of ourselves as analysts according to the rules of logic. Therefore, the production of proof that possibility and conceivability do not coincide does not show that possibility and conceivability cannot

---

[9]Interestingly, Lewis who holds that worlds actually exist argues that there cannot be such a thing as an inconsistent world, although he may mean by this that a world cannot be internally inconsistent even though it might seem inconsistent to us with respect to our logic.

coincide. In other words, humans are not able to achieve the appropriate perspective from which to evaluate the relation between logic and human psychology.

We conclude that the division between what is logically possible and what is logically impossible *is* arbitrarily determined by the contingent features of human psychology. Clearly, this conclusion would cause no difficulty in a theory which held that possible worlds were abstract entities of our own devising anyway. It might appear, however, that it would be a considerable problem for an approach which embraces modal realism, since if possible worlds actually exist then they must exist independent of any human knowledge of or beliefs about them, and so the limits of human conceivability might well not coincide with the boundary between possible and impossible worlds. It may be the case that there are no worlds which contradict the laws of logic which hold in w_0, but it may equally be the case that there are worlds which from our point of view would be logically inconsistent; for example, worlds with mutually contradictory pairs of properties.

In fact, our conclusion does not hamper our acceptance of worlds as existent entities. We can propose that an infinite number of worlds, which we would label logically impossible, actually exist in a modal realistic sense. However, since natural languages function according to the laws of logic as we know them (e.g. entailment relies on the rule of the excluded middle), the worlds which are logically impossible are worlds which we cannot use our languages, in their current states, to talk about, and hence these worlds are irrelevant from the point of view of linguistics at least. In fact, such a world might be classified as irrelevant in an even more fundamental way since any world in which modus ponens does not hold is one which humans would never be able to understand. There is no reason to assume however, that because a world is linguistically irrelevant or incomprehensible to us, it follows that it is actually impossible. Lewis (1986) notes without any anxiety that worlds can outstrip our capacity to describe them, and while this is a strong argument against the adequacy of a "linguistic ersatz" approach it presents no problem for modal realism.

Since we have no reliable way to distinguish between possible and impossible worlds we cannot assume that worlds which we currently believe to be logically impossible are not

contained in W. We cannot be sure that W does not contain distant future counterparts of w_0 which are indescribable with our current languages. Since all of our laws of logic as well as our physical laws are based on observation of the current w_0, (and its earlier temporal counterparts) there is not and cannot be any evidence available to us that our current intuitions about levels of possibility are mistaken. What we can be sure of is that if we attempt to make statements about logically impossible worlds, these statements will be ones to which we cannot assign truth-values. On a practical level then, it does not matter that there may be future counterparts of the current w_0 which are indescribable by virtue of being logically impossible, since we have already accepted that, in general, we have do not have sufficient access to future counterparts of w_0 to assign truth-values to statements about them.

We can conclude that there is only one set of successive temporal counterparts in each W, but that we do not know which particular W we are in. We can identify some close temporal counterparts in the past and we can infer properties of more remote temporal counterparts in the past since we know that they must be consistent with the current w_0. We can predict some properties of *close* temporal counterparts in the future fairly reliably, given that adjacent counterparts differ from their immediate neighbours by small numbers of properties. However, we cannot set boundaries which constrain the very remote future temporal counterparts in any way at all since our division of the set of worlds according to possibility is epistemically based. There are no particular repercussions, for the location theory presented here, from our decision to view differences in world type as epistemically based, since because w_0 shifts according to the location of a discourse, the most possible world is always the world in which some discourse participant exists, and the only accessible temporal counterparts must be consistent with it.

5.3.1 Underspecification.

One important question relating to the bounds of logical possibility concerns underspecified or incomplete worlds. As discussed in section 2.3. of Chapter 2 Fodor (1979) appeals to the notion of an underspecified or incomplete world to account for truth-valuelessness. She

argues that a truth-valueless sentence is about some world where it really is the case that some property neither holds nor fails to hold[10]. For instance, Fodor describes the world where the King of France exists as a "particularly thin fiction" which is not specified for any property save the existence in it of the King of France. She assumes without any discussion that fictional worlds in general are underspecified, merely giving a pair of example sentences to illustrate her point. However, these sentences do not justify Fodor's conclusion that fictional worlds are underspecified. She compares the sentences given here as 5.09 and 5.10.

5.09. Winnie-the-Pooh likes Tuesdays.

5.10. The King of France likes Tuesdays.

Fodor proposes that we do not know the truth-value of 5.09 because the world of Winnie-the-Pooh is not specified for his liking Tuesdays, and that this constitutes proof that the world is incomplete. She argues further that we can generalise this explanation to account for the perceived truth-valuelessness of 5.10. However, the worlds evoked by these two sentences are quite clearly very different in terms of potential possibility with respect to the current w_0. The world in which we locate the referent of the name **Winnie-the-Pooh** in 5.09, which we will call w_0''' is of necessity not a physically possible world since, in several ways, it is inconsistent with what we take to be the physical laws of the real world[11]. Some world w_0'' in which we locate the referent of **the King of France**, on the other hand, could be a physically possible world with respect to w_0. Since w_0''' is a physically impossible world our argument is that, based on physical facts about w_0, we are less able to make inferences about w_0''' than inferences about any worlds w_0'' or w_0', which intervene between w_0 and w_0''' in terms of possibility. That is, we cannot, using what we know about our own cognitive states, make inferences about the extent of, or limits to, the cognitive states of talking teddy bears, which are as reliable as inferences we could make about the cognitive states of individuals such as the King of France; who can exist in physically possible worlds and are presumably human. We are free to choose whether Winnie-the-Pooh is human-like with respect to Tuesdays, or

---

[10] see also chapter 4, section 4.4.2.

[11] For example, if the process of evolution by natural selection operates according to a physical law then the evolution of a teddy bear with the capacity to learn a human language is not a feature of any physically possible world.

teddy-bear-like with respect to Tuesdays. If we choose the former it must either be true or false of him (in every logically possible world) that he likes Tuesdays; if we choose the latter liking Tuesdays is not a property with respect to which he has either a positive or a negative value. Neither decision forces us to admit that w_0''' is underspecified or incomplete; just as the failure of Sherlock Holmes to exist in the world of Star Trek does not render the world of Star Trek incomplete, and the lack of the ability to photosynthesise either by C4 metabolism or by C3 metabolism on the part of humans in the current w_0 does not render this world incomplete.

However, we cannot conclude, simply on the basis that this world does not *have* to be incomplete, that no world *is* incomplete. Clearly, a world which contains entities which do not exist in it, or properties which do not hold of it, is not a consistent world from the point of view of the logical laws of the current w_0; but since we have been forced to accept that our logical laws may not define all worlds, we have to acknowledge that such incomplete worlds may exist. Furthermore, we might be tempted to say, that given the existence of underspecified worlds, we have a ready-made explanation for truth-valuelessness. If there are worlds such that any statement about them is truth-valueless and there are sentences which are intuitively truth-valuelessness, it might seem that associating the two would provide an adequate explanation.

There are two objections to this proposal. The first was made in Chapter 4 section 4.3 where we argued that it is not necessary to attempt to associate truth-valueless sentences with underspecified worlds since we can readily account for their truth-valuelessness on the basis of the limited knowledge of the hearers who attempt to assign the truth-values. The second objection relies on an acceptance of the view of interpretation laid out in the chapter 6. There we argue that the interpretation of a referring expression can be seen as the process of locating its referent and that this process is separate from and prior to the fixing of a truth-value to the sentence in which the referring expression occurs. If we accept this view of interpretation it is nonsensical to argue that the reference world is chosen on the basis that it will prevent the assignment of a truth-value. If the resulting truth-value were an influencing factor in the choice of reference world it would be more plausible that a hearer would choose a reference

world which rendered the sentence true. If the eventual truth-value is not an influencing factor in the choice of reference world, then any proponent of the argument that truth-valueless sentences are about under-specified worlds would have to advance some other kind of justification for this choice. In the following chapter, we set out an explanation of the rules governing the choice of reference world which ensure the very opposite; that an underspecified world is never chosen over any other potential reference world.

5.4 Familiarity

The preceding discussion of counterparts and temporal counterparts does not account for all of the putative different types of worlds which were presented in Table 3.3. in chapter 3 section 3.1.4 where we also distinguished fictional and mythological worlds from other types of counterpart worlds. Such worlds are distinct from the rest of the set of possible worlds in a way which we can define in terms of familiarity. The distinction between familiar and unfamiliar worlds is clearly fundamentally different from distinctions on the basis of possibility or W membership. It was this implicit assumption (that possibility ranking is distinct from any fictional/mythological status) which allowed us to use fictional worlds as examples of worlds at varying levels of possibility. While we concluded that it *may* be the case that our divisions along a scale of possibility have no basis in reality, but are simply reflections of our epistemic state, a division of the set of worlds into familiar and unfamiliar sub-sets is *necessarily* epistemically based since the partition is different for each individual. Let us assume that a fictional world is the clearest example of a familiar, established world that is not w_0 and a world invoked by a counterfactual modal statement to be the best example of an unfamiliar world. The worlds in which we locate the referents of the subject NPs in 5.11 and 5.12, for example, can be taken as representatives of worlds of these two types.

5.11. Scarlett O'Hara was not beautiful.

5.12. Chelsea Clinton's sister might be called Belgravia.

Although the amount of knowledge we have about these two worlds is different and our potential for assigning truth-values to statements about individuals in each of them is correspondingly different, there is no justification, if we accept that worlds exist in a modal realistic sense, for arguing that the worlds themselves are of fundamentally different types from each other. There is no property or set of properties for some world which can determine whether that world is a fictional world. Fictional worlds are simply those counterparts with which we are familiar from works of fiction. To illustrate what we mean by familiarity, let us consider sentences 5.13 and 5.14.

5.13. Scarlet O'Hara is a Southern Belle.

5.14. Belgravia Clinton is a Bloomsbury bluestocking.

5.13 can be interpreted as a true statement about an individual in a familiar, fictional world, the world invoked by *Gone with the Wind*, which we will call w_0", while 5.14 can only be interpreted as a true statement if the subject NP is interpreted as referring to an entity in a different counterpart world which we will call w_0"'[12]. In terms of their inherent properties these two worlds have to be taken to be of exactly the same type. Both worlds are connected to the current w_0 by a number of counterpart links. That is, starting from the current w_0, it is possible to change the value of one property to arrive at a one-place counterpart of w_0 and then to change a second property of this world to arrive at a two-place counterpart of w_0 and so on, until we arrive at a world in which either every proposition asserted in *Gone with the Wind* is true, or it is true that a counterpart of President Clinton has another daughter who is called Belgravia and who is a Bloomsbury bluestocking.

However, we want to be able to distinguish these two worlds to reflect the intuition that one, w_0", is an established familiar world which we have accessed before and will presumably access again, and one, w_0"', is accessed for the first time for the purpose of interpreting the current sentence and may never be accessed again. To do this we propose a final relation,

[12]Note that all we are saying here is that these worlds are potential reference worlds which will produce truth-values of true. We are not arguing that these sentences are perceived as true or predicting that these reference worlds would be chosen by a hearer in a discourse. The issue of how reference worlds are chosen is the topic of the next chapter.

which we will call an epistemic link R_e, which will define another layer of structure on the set of possible worlds by which familiar worlds are distinguished from counterparts in general.

An epistemic link R_e connects an individual $i \in A$ in w_0 to w_0' if w_0' is a familiar world with respect to i. A familiar world for i, is a world w_0' in which at least one $a \in A'$ has at least one property for which i knows the value. That is, w_0' is a familiar world for i, if i has assigned a truth-value to some sentence which expresses a proposition about an entity in w_0'. Because we propose that the location of referring expressions precedes the assignment of truth-values during interpretation, there must be some previous sentence which a hearer has assigned a truth-value to that establishes a world as familiar during the interpretation of a current sentence. This distinguishes the world w_0'' in which Scarlet O'Hara exists from a world w_0''' in which Belgravia Clinton exists since we are familiar with the value of *further* properties of w_0'' besides those which are specified in sentence 5.13[13]. Since only one assignment of a truth-value is necessary to establish a world as familiar, the set of familiar worlds contains many more worlds than the generally familiar worlds of fiction and mythology. The set of familiar worlds for some individual will contain, as well as these which are familiar worlds for many other individuals, counterparts invoked by private jokes which are only familiar worlds for a restricted number of individuals, and dream-worlds or fantasy worlds which might be undistinguished counterpart worlds for every other individual.

We will not distinguish between the initial act which connects w_0 through some individual to some w_0' for the first time, from any subsequent act which establishes that w_0' as a familiar world for other individuals. Thus, we are forced to assume that the author of some fiction has the same type of epistemic link with the fictional world as any reader of that fiction. This is slightly at odds with the view that an epistemic link is brought about by an act of truth-value assignment, since if a work of fiction is seen as a discourse, the author is the speaker and the reader is the hearer, and we have, up until now, characterised truth-value assignment as a process carried out by the hearer in a discourse. However, since there is no possibility of de

---

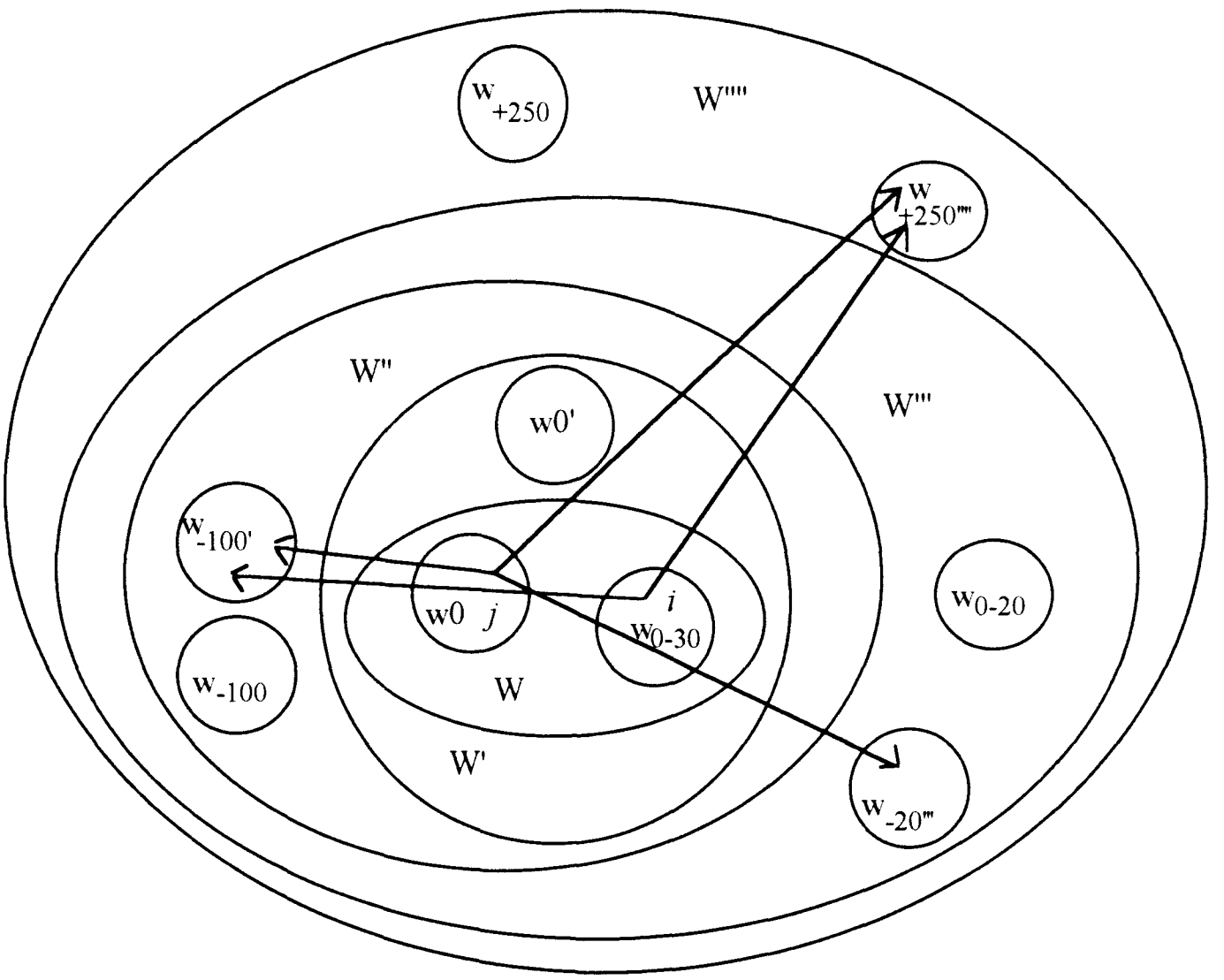
[13]If we think of w_0''' as a world exactly like w_0 except for the properties specified by the propositions expressed in 5.14 it might appear that we have, in fact, determined the value of further properties of that world, but it should be remembered that w_0''' is a counterpart of w_0 and although properties of w_0 are *copied into* w_0''' the sets of items and attributes of w_0''' are not the same entities as the sets of items and attributes in w_0.

re knowledge about fictional worlds, we can argue that, given a modal realistic view of worlds, both an author and a reader of a piece of fiction, when they assign truth-values in the fictional world, are *choosing* some world from among the set of existing worlds. The argument that an author creates a world, which a reader only views, is at odds with the fundamental idea of modal realism that all worlds actually exist.

Figure 5.6. shows a partial diagram of the set of epistemic links in U.

Figure 5.6.

Epistemic links in Log. P



In this diagram w_0 is the current world, W is the set of temporal counterparts of w_0 and $w_{0\text{-}30}$ is the past temporal counterpart of w_0 1964. W'' is the set of technologically possible counterparts of w_0 and in it $w_{\text{-}100}'$ is a world containing Sherlock Holmes and $w_{0\text{-}100}$ is a

world containing Sherlock Holmes and his wife. W''' is the set of physically possible counterparts of w_0 and in it w_{-20}''' is a world containing Col. Steve Austin, w_{-20} is a world containing Col. Steve Austin and his twin daughters. W'''' is the set of logically possible counterparts of w_0, and in it w_{+250}'''' is a world containing Mr Chekhov, and w_{+250} is a world containing Mr Chekhov and his prize-winning guinea pig. There is an entity *i* which bears the name **Gene Roddenberry** in w_{-30}, the temporal counterpart of w_0 which held in 1964 and an entity, *j* which bears the name **Norma Major** in w_0.

If the single headed arrows in Fig. 5.6. are read as representations of epistemic links, then we can understand the following relations from the depicted links.

There is an initial epistemic link R_e^i from *i* in w_{-30} to a world w_{+250}'''' in which an entity which bears the name **Mr Chekhov** exists which (along with many other properties) identifies w_{+250}'''' as the world of Star Trek. There is no such epistemic link between *i* and w_{+250}. We can assume, on the strength of the cultural importance of Sherlock Holmes, that there is also an epistemic link R_e^i between *i* in w_{-30}, and w_{-100}', the world invoked in the stories of Arthur Conan Doyle, but no such link between *i* and w_{-100}. We also assume that there is an epistemic link from *j* in w_0 to w_{-20}''; i.e. we assume that Norma Major has assigned truth-values to statements about entities in the world invoked by The Six Million Dollar Man. In contrast, there is no possible epistemic link between *i* in w_{-30} and w_{-20}'', since the initial epistemic link from an individual in W is from a temporal counterpart which is not accessible from w_{-30}, that is from a future temporal counterpart.

There is one final stipulation we must make concerning the structure of the set of possible worlds before moving on to a discussion of the location theory itself. As we noted above, any familiar world which is accessible to an individual in some other world by an epistemic link, is also accessible via a string of counterpart relations between the two worlds. Since an epistemic link is a *direct* link from some individual to another world and, in contrast, the n-place counterpart relation which connects two worlds is an indirect link via n-1 intervening counterparts, we can say that a familiar world such as w_{+250}'''' is closer to w_0 than its close counterparts such as w_{+250}, if we take epistemic links to be *short-cuts* to other worlds.

5.5. Summary.

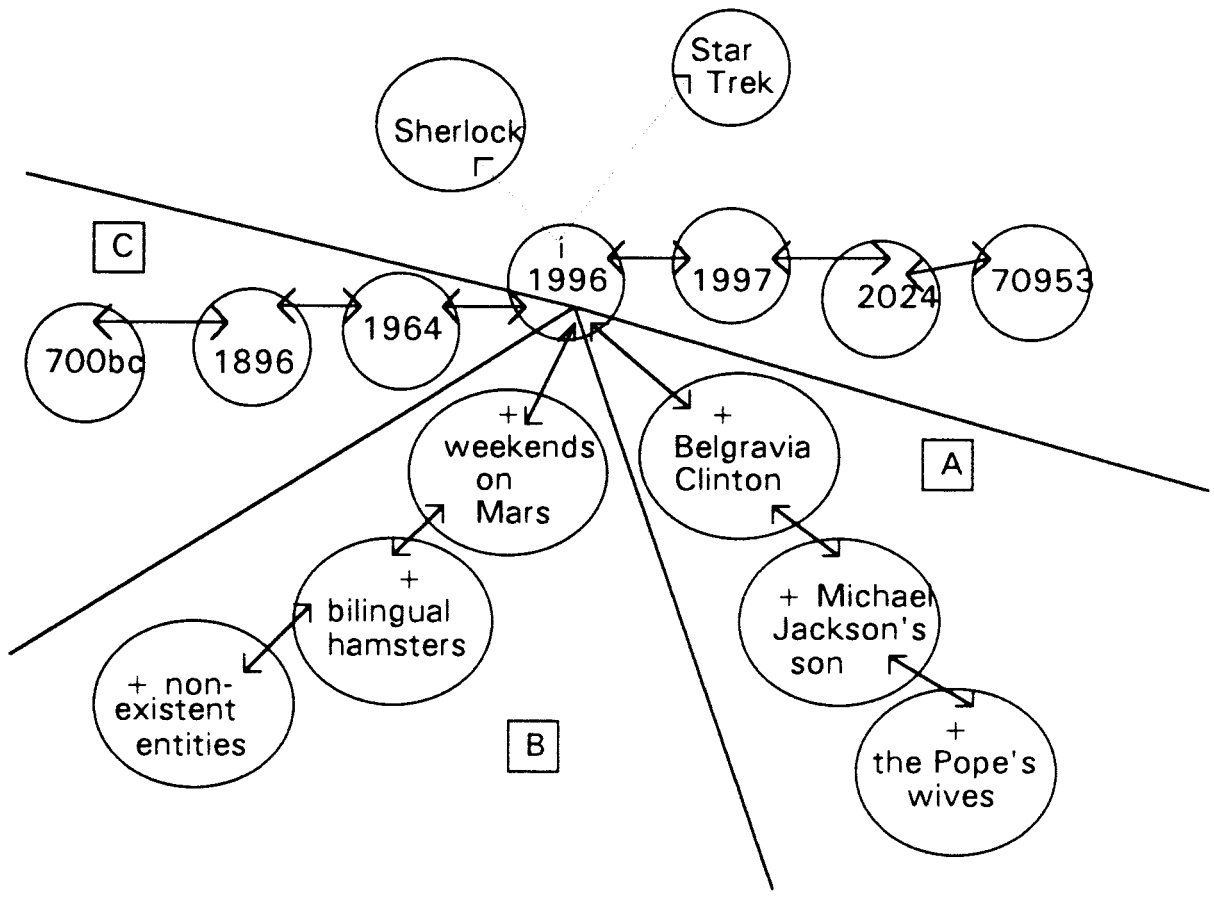
In this chapter we have established a complex, structure-rich set of worlds arranged across several dimensions according to a principle of relative proximity. We proposed that given some world w_0, a set of counterpart relations can be defined which connect all other worlds to w_0, such that across an infinite number of dimensions the nearest world to w_0 is an almost identical copy of it which differs from it by the value of one property or minimal bundle of properties, the next nearest by two properties or minimal bundles of properties and so on, where a property of a world is the existence in it of some item or attribute. We discussed the possibility of defining a particular set of counterparts which are temporally consistent; having acknowledged the difficulty of stating the definition of such a set in a non-circular way, we opted to depend on the evidence of clear cases, to define the distinction between consistent and inconsistent sets of worlds. Finally, we proposed sets of epistemic links which hold between individuals and worlds with which these individuals are familiar; familiarity being defined as the relation a hearer has with a world in which he or she has assigned a truth-value[14].

Fig. 5.7. represents a partial diagram of some of the counterparts of the current w_0 (labelled 1996).

[14]A hearer has assigned a truth-value "in" a world if there is an entity in that world and a statement which the hearer has interpreted as being about that entity, and the hearer has assigned a truth-value to that statement.

Figure 5.7.

Three dimensions of the set of possible worlds.



A contains worlds increasingly remote from w_0 with respect to their ontology: in the nearest world, the counterpart of Chelsea Clinton has a sister; in the next nearest the counterpart of Chelsea Clinton has a sister and the counterpart of Michael Jackson has a son; and another more remote world contains counterparts of these two individuals and individuals which are the wives of a counterpart of the Pope. **B** represents a series of counterparts of w_0 which are increasingly remote with respect to their possibility according to the properties of w_0: a world where weekends on Mars are available is technologically impossible, a world where bilingual hamsters exists is physically impossible, and a world containing non-existent entities is logically impossible. **C** represents some of the temporal counterparts of w_0. Note that those worlds which may be the future versions of w_0 are not distinguished from other types of counterparts. The dotted lines in fig. 5.7. represent epistemic links between some individual in 1996 and two worlds where the named entities exist.

In the following chapter we will employ this structured set of worlds in a discussion of problems surrounding the interpretation of referring expressions and the assignment of truth-values to utterances containing them.

Chapter 6.
Interpretation Using Possible Worlds

6.1. Interpretation

The theory we propose is one which applies to discourses rather than isolated sentences, embraces pragmatic as well as semantic aspects of meaning and takes account of the epistemic states and limits of discourse participants. In this framework we use the term "interpretation" to refer to a dynamic process by which the hearer in a discourse evaluates utterances of both referring expressions and sentences. The discussion of interpretation in this chapter depends on three assumptions which we will attempt briefly to justify: we assume; a. that the interpretation of referring expressions is fundamentally different from the interpretation of statements about them; b. that the interpretation of referring expressions can be seen as the act of locating referents while the interpretation of statements can be equated with truth-value assignment; c. that truth-values are properties of utterances.

6.1.1. Distinguishing the Interpretation of Referring Expressions from the Interpretation of Utterances.

One important distinction between a location theory and a semantics which employs a formal model theory (such as DRT or Montague Grammar; Kamp and Reyle, 1993; Cann, 1993) is that instead of embedding whole sentences in worlds (or models), a location theory embeds individual referents in worlds. As a result we can analyse sentences as being about individuals, rather than about worlds. We are following Frege in this who maintains that sentences are not about worlds. He stated that "The *referents* of our worlds are what we talk about." (Geach and Black, 1980). It is this which allows us to analyse sentences with pairs of referring expressions as statements about individuals in two or more separate worlds, such as sentence 6.01 and 6.02, as well as sentences with pairs of referring expressions which co-exist in the same world, such as sentence 6.03.

6.01 The Tin Man and Dr Watson called on J.R. Ewing for tea.

6.02 Oliver Sacks looks a lot like Santa Claus.

6.03 Oliver Sacks looks nothing like Jeremy Paxman.

By advocating a split between the interpretation of referring expressions and the interpretation of whole sentences, we can handle so-called presuppositional effects. Let us consider, sentence 6.04, for example:

6.04 Somewhere a student is weeping.

The interpretation component of DRT requires a hearer of 6.04 to find some individual and check that that individual is first, a student and second, weeping. In contrast, a location theory requires a hearer of 6.04 to find some individual that is a student and check that they are weeping. This distinction can buy us the possibility of arguing that *where* we find a referent has a different effect on the interpretation of a sentence than *whether* we find a property.

6.1.2. Modelling Interpretation.

The second decision we must justify is our decision to equate the interpretation of whole utterances with the act of assigning truth-values.[1] It is widely accepted in the fields of semantics and pragmatics that communication is not primarily concerned with the transferring of knowledge of true statements from speaker to hearer, and that correspondingly truth-value assignment is not equivalent to interpretation. Lyons states that ". . . the transmission of descriptive information is not usually an end in itself." (Lyons, 1977: p725) and claims that ". . . the sender's desire to convince the receiver that such-and-such is true commonly derives from or is associated with some other purpose." (ibid.: 34) He proceeds to make a tentative suggestion that " . . . it may be the case (and it commonly is) that what we actually say is

[1]Clearly, we must also justify taking the interpretation of referring expressions to be an act of location. However, this cannot be done until the apparatus of the current theory is in place. The working of the location theory is the topic of much of the remainder of this chapter.

uninformative, in that the receiver knows (and we know that he knows) whatever fact it is that we are drawing to his attention." (ibid.: 35).

Sperber and Wilson (1986) go even further towards de-emphasising truth-value assignment in the area of interpretation as a whole by suggesting that not only is truth or falsity not central to linguistic communication but that linguistic communication is not central to communication as a whole: "we believe that the kind of explicit communication that can be achieved by the use of language is not a typical but a limiting case." (Sperber & Wilson, 1986: p55). Sperber and Wilson explain the recurrent interest in the role of truth in communication as an artefact of the limits of linguists, although it is not clear whether they mean to imply that linguists in general are lazy or simply unimaginative.

> "With assertions, often taken to be the most basic case, the informative intention is taken as the intention to induce in an audience the belief that a certain proposition is true. There is a very good reason for anyone who is concerned with the role of inference in communication to assume that what is communicated is propositional: it is relatively easy to say what propositions are and how inferences might operate over propositions."
>
> (Sperber & Wilson, 1986: p57)

What then is our justification for concentrating, not only on the interpretation of declarative sentential linguistic acts of communication but restricting our attention to the reputedly small and unimportant part of interpretation that is truth-value assignment? We are not convinced that communicative acts which rely on the transmitting and decoding of factual statements are as rare as the quotations above suggest. This thesis, for instance, is an example of an extended act of communication which is entirely linguistic and is to a large extent, concerned with transmitting factual statements. Few linguists would claim that communication generally consists *only* of transmitting and assigning truth-values to assertions, but on the other hand few would deny that these procedures are not part of linguistic communication. We claim that interpretation using the system of implicatures introduced by Grice (1975) can only be initiated once it has been established by a hearer that an utterance is false, since it is this which allows the inference that the factual statement it expresses cannot be the information intended by the speaker to be recovered (Grice, 1975). Lyons also concludes that ". . . it may be assumed, however, that the interpretation of non-informative utterances trades upon our

ability to interpret the same utterances in contexts in which they would be informative." (Lyons, 1977: 35). and Levinson includes "truth" as the first component of the communicative content of an utterance (Levinson, 1983). If the assumption is correct that truth-value assignment is an important part, although perhaps not the ultimate goal, of interpretation, then it is crucial to have a satisfactory explanation of the rules governing the process.

6.1.3. The locus of truth.

The final assumption we must defend is that truth-values are properties associated with utterances. There are two other types of linguistic item which might be regarded as the bearers of truth-values; propositions and declarative sentences. The choice between the three options, is perhaps more a matter of terminology than anything more fundamental since, as we claim below, the proponents of the argument that truth-values are properties of propositions, and the proponents of the argument that truth-values are properties of sentences, generally acknowledge the role of utterances as well.

There are two difficulties, although as we noted they may not be serious ones, which arise from taking a proposition to be the locus of a truth-value. First, propositions are notoriously difficult to define in a satisfactory way. Propositions have been variously defined as, concepts, meanings, parts of meanings, and the sense of sentences; all definitions which themselves rest on ambiguous terms, or at least terms which are difficult to define. It is a contested point, besides, what *kind* of things propositions are, whether psychological entities or abstract constructs (Lyons, 1977). The association of truth with propositions when propositions themselves have no universally accepted definition leads to the appearance of complete mutual incompatibility between different theories concerned with truth. For example, consider the following quotations from Cann (1993) and Bradley & Swartz (1979):

> "A proposition may be true or false." (Cann, 1993: 14)
>
> "Propositions are those things to which truth and falsity may be attributed." (Bradley & Swartz, 1979: 79)
>
> "Propositions may vary in their truth value from utterance occasion to utterance occasion" (Cann, 1993: 14)
>
> "We shall assume . . . that propositions are omni-temporally (and for that matter omni-spatially) true or false." (Bradley and Swartz, 1979: 107)

The underlying cause of such apparent incompatibility is, we claim, bound up with the second problem with associating truth with propositions, which is that it is inarguable that truth or falsity is tied to utterance occasions; but because there is no accepted definition of a proposition, there is no accepted way to associate propositions with utterance occasions. This allows Bradley and Swartz to assume that identical utterances produced on two separate occasions with opposing truth-values must be associated with two different propositions, but it also allows Cann to assume when identical utterances are produced on two separate occasions with two different truth-values, that they are associated with the same proposition, but that the proposition is associated with a different truth-value on each occasion. It is worth noting that Bradley & Swartz, who claim that truth-values (associated with propositions) are timeless, are forced as a result to distinguish between "truth" and "actual truth" and they concede that in deciding whether or not truth refers to actual truth "what matters is *who* uses it [i.e. the term *true*]" (Bradley & Swartz, 1979).

Some formal linguists working in Montague Grammar or other theories which employ predicate logic (e.g. DRT), associate truth and falsity with sentences (or formulae or grammatical strings) (Dowty, Wall and Peters, 1981). Again, it is not clear whether the association of truth with sentences in Montague Grammar constitutes a serious theoretical claim which is distinct from associating truth with propositions, or is simply a convenient use of terminology based on the assumption that sentences and propositions co-occur. In addition, since "most sentences in natural language are context-dependent in some way or other" (Allwood , Andersson & Dahl, 1977:)[2] it appears that the sentence which is assigned a truth-value in Montague Grammar is understood to be a sentence with a context attached.

---

[2]In fact, if tense is viewed as a type of deixis then all sentences are context dependent.

It appears then that both theories which associate truth with sentences, and theories which associate truth with propositions, appeal to features of the context. We can conclude that if it is possible in the framework of a theory to associate truth with utterances (which unite sentences, propositions and contexts) then there is no fundamental objection to proposing this particular option. The theory of interpretation presented in this thesis takes a dynamic, discourse level approach in which the roles of speaker and hearer are differentiated; and truth-value assignment is understood as a process undertaken by specific hearers. Because of this the current framework is particularly well-suited to an analysis of truth as a feature of utterances[3].

6.2. Interpretation Using Possible Worlds.

The central problem which we have identified in combining a modal realistic view of possible worlds and an analysis of the interpretation of linguistic expressions, is that, if we accept the claims of modal realism, we must accept that for any statement, there are worlds in which that statement is true and worlds in which it is false, since the set of worlds contains every possible combination of entities, properties and relations. However, since not every statement uttered is perceived as true (or false), predicting truth-value assignment is equivalent to predicting the choice of reference world. We must now show that the proposed arrangement of worlds translates into some means by which the choice of reference world made during interpretation can be predicted.

6.2.1. Accessing the Current World.

The first proposal we make is that world choice is dictated in a very simple way according to

---

[3]It is perhaps unfortunate, given our claim that truth-values are properties of utterances, that the present written format forces us to talk about the truth-values of written tokens of sentences. We persevere with this however, since it is only if readers evaluate competing potential interpretations with respect to the real, current, context of this thesis that we can avoid the effects of any implicit filling-in of some other context which may vary from reader to reader, and which might affect intuitive truth-value assignments in a potentially misleading way.

accessibility; and that the most accessible world is the nearest world. That is, hearers will try to locate referents for referring expressions in the nearest world along some dimension. Since discourses take place in worlds, we predict that hearers will try to locate referents in the world in which the discourse is taking place. If this idea is correct we would predict that 6.05 in the current w_0, where the nearest world is the world in which this thesis is being written, will be interpreted as a reference to an entity in this world, and that 6.06 will be interpreted as being a statement about that entity. 6.06 will be assigned the truth-value *true* if it is considered to represent an accurate statement about that entity.

6.05. Richard Branson.

6.06. Richard Branson has done rather well for himself.

To be sure that the choice of reference world for 6.05 and 6.06 is determined by accessibility alone and is not influenced by any pressure to interpret the sentence as true, we must determine what reference world is chosen in the interpretation of sentences which we perceive to be false and sentences which we perceive to be truth-valueless. If we consider sentences 6.07 and 6.08, we can see that these too are perceived to be statements about an entity in the current w_0, even though our intuition may not assign them the truth-value *true*.

6.07. Richard Branson is poor but proud.

6.08. Richard Branson sleeps soundly

If the choice of reference world for 6.07 was influenced by some kind of pressure to interpret the sentence as true, it would be expected that 6.07 would be interpreted as a statement about an entity in a counterpart of w_0 in which the counterpart of Richard Branson was poor but proud. This does not appear to be the case. Instead we interpret 6.07 as a false statement about an entity in w_0. Likewise, if the choice of reference world was governed to some extent by the pressure to assign *some* truth-value, no matter which one, then we would predict that a hearer of sentence 6.08 who did not know whether or not Richard Branson in w_0 sleeps soundly, would interpret sentence 6.08 as a statement about an entity in a counterpart world. Again, the intuition about 6.08 does not support this analysis; it is perceived to be a statement

about an entity in w_0 to which we cannot assign a truth-value. On the strength of these three examples, we can state, as a rule of interpretation, that hearers will locate referents in w_0, no matter what truth-value, if any, this yields.

6.2.2. Accessing Familiar Worlds.

Recall that we suggested that epistemic links between an individual in one world and some other world should be understood as short-cuts between worlds. This allows us to retain the characterisation of accessibility as equivalent to proximity and to define some individual's set of familiar worlds as the nearest and hence the most accessible worlds after w_0 (the world in which the individual exists). On the strength of this we predict that 6.09 and 6.10 are interpreted as statements about two separate worlds, which are both very remote from w_0 in terms of counterpart relations, but are accessible to many individuals in w_0 via direct epistemic links.

6.09. Scarlet O'Hara was not beautiful.

6.10. Sherlock nestled in silence into his heavy coat.

It should be noted that although 6.09 and 6.10 are identical to sentences which appear in *Gone with the Wind* and *The Hound of the Baskervilles*, it is not their resulting indisputable truth which influences the choice of reference world, since sentences 6.09a and 6.09b are interpreted as being about the same entity as 6.09, and 6.10a and 6.10b are interpreted as being about the same entity as 6.10 even though the a. sentences are perceived as false and the b. sentences cannot be assigned any truth-value.

6.09a. Scarlet O'Hara was a Yankee carpetbagger.

6.09b. Scarlet O'Hara never scratched her chicken pox.

6.10a. Sherlock Holmes was kind but stupid.

6.10b. Sherlock Holmes rarely changed his socks.

By accepting that an epistemic link is a short-cut we can continue to think of the world invoked by *Gone with the Wind* as nearer or more accessible to w_0 than some counterpart world which is identical to the w_0 except for the existence in it of Scarlet O'Hara, even though this counterpart has more properties identical to properties of w_0 than the world invoked by *Gone with the Wind* has[4].

6.2.3. Accessing Counterparts.

6.2.3.1. Sign-posted Counterparts.

The remaining worlds available as the locations of referents are those which we labelled counterpart worlds. We will make a distinction between two types of counterparts on the basis of the processes by which they are accessed. The first of these, we will call *sign-posted counterparts* which may be accessed when syntactic constructions act as sign-posts that indicate the need for shifts into counterpart worlds. The syntactic sign-posts which trigger shifts into counterpart worlds are modal constructions such as *would* and *might* and the counterfactual *if ... then* construction. 6.11 - 6.13 are examples of sentences which are interpreted by locating the subject NPs in sign-posted counterparts.

6.11. Camille Paglia's daughter would never play with Barbie dolls.

6.12. Chelsea Clinton's sister might not be a Democrat.

6.13. If the Pope's wife were a Catholic, then the Cardinals would be happy.

We propose that *would*, *might* and counterfactual *if ... then* in 6.11 - 6.13 indicate the need for a shift into a *minimal* counterpart world of w_0. A minimal counterpart world of w_0, for 6.12 for example, is a world identical to w_0 except for the existence in it of an individual who is the sister of the counterpart of Chelsea Clinton in w_0.

---

[4]If the propositions explicitly stipulated in *Gone with the Wind* were analysed as a minimal bundle of properties associated with the existence in any world of Scarlet O'Hara, then the familiar world would be a minimal counterpart of w_0. However, this cannot be correct. If all the properties of worlds are copied into minimal counterparts, then it would not be possible to make modal statements.

The reference worlds employed in the interpretation of sentences 6.11 - 6.13 are counterparts on two counts since even without the syntactic signposts, the referents of the subject could not be located in familiar worlds. However, we propose that syntactic signposts alone can trigger the location of a referent in a counterpart, e.g. sentences 6.14 - 6.16:

6.14. The Princess of Wales would never wear a nose ring.

6.15. The Rolling Stones might retire some day.

6.16. If Ian Botham were ten years younger, England could win the next Test.

We claim that syntactic signposts also facilitate moves into counterparts of other established worlds as well as w_0. This is shown in examples 6.17- 6.19.

6.17. If Scarlet O'Hara had been a Yankee carpet-bagger she would never have been hungry or poor.

6.18. If Sherlock Holmes was stupid, he might be happier and easier to live with.

6.19. If Mr Spock's wife were a Klingon, their children would not be human.

Again, we emphasise that the choice of reference world for the type of sentence currently under discussion is independent from the question of truth-value assignment. Some sentences interpreted using sign-posted counterparts are perceived to be true, e.g. 6.20, some are perceived to be false, e.g. 6.21; and some cannot be assigned a truth-value, e.g. 6.22.

6.20. Chelsea Clinton's sister might not be a Democrat.

6.21. Sherlock Holmes' brother could not have a birth-mark on his bottom.

6.22. If the Pope's wife were a Catholic, the cardinals would be happy.

The traditional role of possible worlds in modal logic has been to facilitate an account of the truth-values of modal and counterfactual statements (McCawley, 1981). The set of possible worlds as a whole has been invoked, along with such additional theoretical tools as Lewis's operation of local entailment (Lewis, 1973), and Nute's notion of a threshold of relevance.

(Nute, 1975), to account for the semantics of modals and counterfactuals in terms of quantification over sub-sets of worlds. It may appear that the use of possible worlds in the current theory, where they are employed individually as locations for referents, results in a net loss of explanatory power since we do not claim to be able to predict a truth-value for every modal sentence.

We claim, instead, that the possibility of assigning a truth-value to a sentence interpreted using sign-posted counterparts rests, as in other cases, on the particular property or relation being predicated of the entity chosen as referent and the hearer's epistemic access to and consequent knowledge about the reference world. The reference worlds employed in the interpretation of 6.20 - 6.22 are unfamiliar worlds and the properties predicated of Chelsea Clinton's sister, Sherlock Holmes' brother, the Pope's wife, and the cardinals are contingent. Therefore the interpreter of 6.20 - 6.22 is unable to assign truth-values just as he or she is unable to assign truth-values to 6.23.

6.23. The Princess of Wales' first nanny had green eyes.

When we look closely at expositions of the meaning of counterfactuals and modal statements which invoke quantification over possible worlds, we typically find only statements of their truth-*conditions* not predictions their truth-*values*. For example, McCawley (1981) discusses the counterfactual sentence repeated here as 6.24.

6.24. If Rosalyn had had an affair with Kennedy, Jimmy would have divorced her.

McCawley (1981) argues that for 6.24 to be true, supposing

> "...that there are three worlds in which Rosalyn had an affair with Kennedy that are the same distance from the real world and are closer to the real world than any other worlds in which Rosalyn had an affair with Kennedy ... he must have divorced her in all three of those worlds."
>
> (McCawley, 1981: p322).

This much is unexceptional. But we claim that for 6.24 to be *perceived* as a true statement on some occasion of utterance would require an impossible degree of access to other worlds in that we would have to be able to verify unknown properties of the minimal counterpart in which Rosalyn had an affair with Kennedy.

In the discussion so far we have only considered counterfactual conditionals. It is our proposal that the interpretation of *indicative* conditionals do not necessarily trigger shifts to counterparts in order to locate referents. Sentence 6.25 for example is about entities located in the current w0.

6.25. If Bill Clinton is incubating diphtheria, then Hilary is too.

It *is* necessary to invoke possible worlds, in order to state the truth-conditions of 6.25. using modal logic, but it is not necessary to invoke an individual sign-posted counterpart of w_0 as the existential location for either of the required referents. Clearly however, it is often the case that at least the consequent clause of an indicative conditional, contains a modal verb and if so then *this* syntactic feature can trigger a shift to a sign-posted counterpart. For example:

6.26. If Bill Clinton is incubating diphtheria, Hilary might die soon.

If both clauses describe future states of affairs then again sign-posted counterparts will be employed as existential locations, but this location will be triggered by the modals not by the indicative construction. For example:

6.27. If Bill Clinton loses the election, Hilary will start using her own name again.

We will now turn to a discussion of the processes involved in accessing temporal counterparts which share features of familiar worlds and sign-posted counterparts.

6.2.3.2. Temporal counterparts.

In chapter 5, we discussed relationships between temporal counterparts in terms of consistency. Another feature of some of the temporal counterparts of a world w_0, which we will focus on in the following discussion of how temporal counterparts are accessed, is that they may be familiar worlds for the individuals who exist in w_0. For example, an individual may have direct experience of a temporal counterpart by virtue of having lived through it, or may have indirect experience of it by virtue of written histories, oral legends, archaeological discoveries etc. We propose that familiar temporal counterparts are accessed in exactly the same direct way as other types of familiar worlds. For example, the referents of the subject NPs of sentences 6.28 and 6.29 are both located via direct epistemic links to familiar worlds.

6.28. Queen Elizabeth I was very fond of Sir Walter Raleigh for a while.

6.29. Gandalf the Grey was very fond of Bilbo Baggins.

The view that the set of temporal counterparts of the current w_0 contains sub-sets of familiar worlds for individuals is made possible by our decision to restrict the application of R_c links which hold between w_0 and other worlds to earlier versions of w_0. Future versions of w_0, we claim, cannot be linked with w_0 via R_c and hence it is not possible for individuals in w_0 to have epistemic links with these worlds.

All sentences which are interpreted as statements about familiar entities in temporal counterparts have an additional feature, however, which is not associated with sentences about entities in other types of familiar worlds. This is that the tense of the verb indicates the temporal link between the reference world and the world in which the discourse takes place. We can think of tense as another type of syntactic signpost. We propose that the direction of the syntactic signpost into the temporal counterpart and the established location of the familiar entity must be compatible, if a sentence is to be interpreted as a non-anomalous statement about an entity in a temporal counterpart.

For example, the subset of temporal counterparts of w_0 which contain the entity Queen

Victoria are familiar worlds for many individuals in the current w_0. Any instance of the referring expression **Queen Victoria**, is therefore interpreted using one of these familiar worlds. If a syntactic signpost, in a sentence containing the referring expression **Queen Victoria**, also points into an earlier temporal counterpart then there is no difficulty. Thus, we predict that sentence 6.30 will be interpreted as a statement about an entity in an earlier temporal counterpart of w_0 and that the reference world is chosen via an epistemic link. The past tense form of the verb merely confirms the location of the reference world with respect to w_0.

6.30. Queen Victoria had eight children.

In contrast, sentence 6.31 is anomalous in the current context.

6.31. ?Queen Victoria has eight children.

The name **Queen Victoria** takes us to an earlier temporal counterpart of w_0 via an epistemic link, but the present tense form of the verb does not confirm this choice. We claim that the reference world is *chosen* on the basis of the referring expression and that in this case the syntactic signpost cannot force a re-analysis, since no referent is available in the current world[5].

The opposite effect can be observed in sentence 6.32, where there *is* a referent for the definite in the current world, but the tense of the verb, which is a syntactic signpost into an earlier counterpart, forces a re-analysis.

6.32. The Prince of Wales, a tragic figure, died young.

---

[5] McCawley (1981) discusses the following pair of sentences which on his theory should be true but are perceived as anomalous (McCawley, 1981: 316):

??if London were in England, the Queen of England would live in London.
??if Nixon had resigned in 1974, Ford would have become president.

It may be that the anomaly lies in the fact that there are referents available in w_0 and yet syntactic signposts point us towards counterparts.

We claim that the re-analysis is due to the availability of a referent for **the Prince of Wales** among the earlier temporal counterparts of w_0 and cannot be explained as the result of pressure to produce a reading which can be interpreted as true, or even one which can be assigned a truth-value. If **the Prince of Wales** is taken to refer in w_0 the sentence must be interpreted as false. We predict that even for hearers who are unable to assign a truth-value to 6.32, if it is taken to refer in an earlier counterpart, will locate the referent there, and that only hearers who did not know that "Prince of Wales" is a hereditary title rather than a name would interpret 6.32 as a false statement about an entity in the current w_0[6].

We can show that re-analysis triggered by a syntactic signpost *can* occur when the original location of a referent is an earlier temporal counterpart, if there is an available referent in the current w_0. Because a referring expression is generally interpreted as referring to an entity in w_0 if one is available anyway, we must set up a context to block *this* location process, in order to see that a syntactic signpost can also trigger location in the current world. Consider sentence 6.33:

6.33. The Madness of George III is set in the time of Parliament's first attempt to impose a regency. Many MPs on both sides of the house say they enjoyed it.

We claim that, given the priming context, **many MPs on both sides of the house** is first interpreted by locating its referents in 1788. The present tense form **say** does not confirm this location and so the indefinite is re-analysed as a reference to a sub-set of the set of MPs in w_0. Again, we cannot argue that truth-value assignment is a factor in this choice as the second sentence (disregarding its tense) is equally plausible on either reading; i.e. as a statement that many Tory and Whig MPs enjoyed the Regency Bill or as a statement that many Tory and Labour MPs enjoyed the play. We conclude that the interpretation of the indefinite is not affected by the hearer's ability or inability to assign a truth-value on either reading.

By proposing that referents are located in ws rather than in Ws we have avoided any

---

[6]see chapter 7 section 7.2.2. for a discussion of the distinction between the Prince of Wales as a name and a referring expression

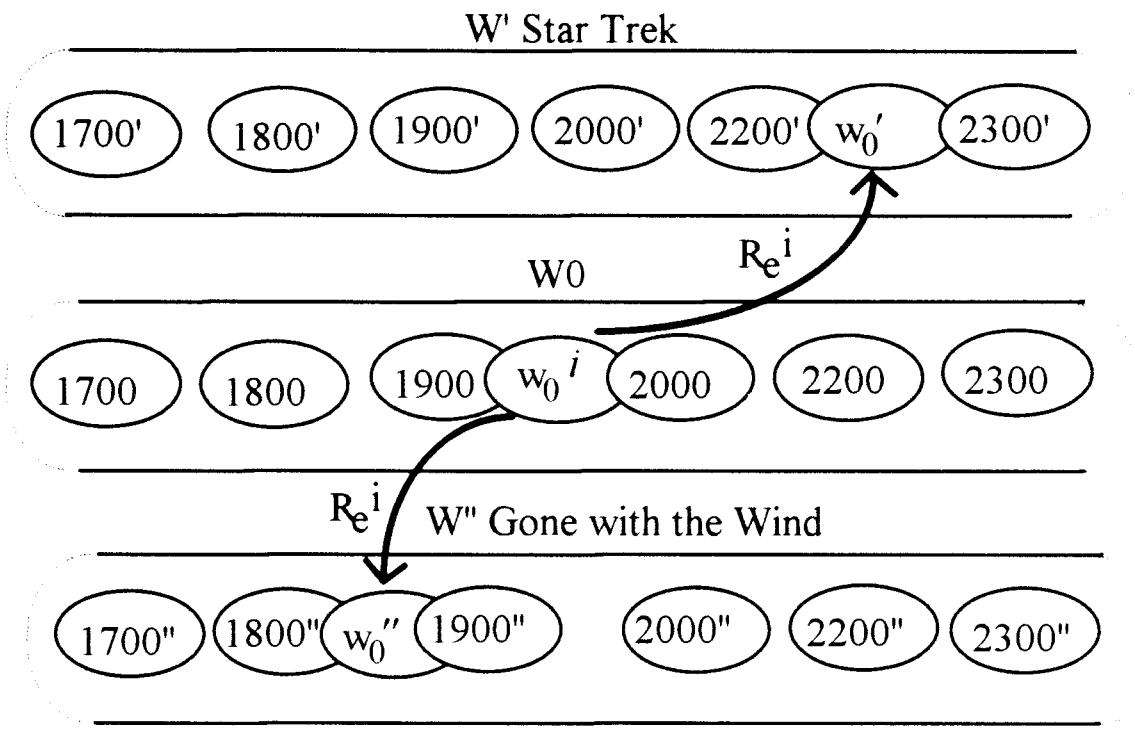
discussion of the implications of the perdurance through time of real-world entities or of the different interpretation processes associated with different verb types; e.g. "instantaneous" events such as *drop a cup* and states such as *be asleep*. These are complex issues and some of the potential problems they pose will be dealt with in chapter 7. However, since some of the difficulty in this area arises from our lack of epistemic access to any future versions of w_0 the same problems do not arise when we consider the operation of syntactic signposts with respect to familiar worlds other than w_0.

There are several factors which might influence the interpretation of syntactic signposts with respect to these worlds. Up until now when we have discussed fictional entities we have referred to their existence in counterparts of w_0; that is other temporally instantaneous worlds. However, since fictions and myths are not references to instantaneous events but describe sequences of events which take place in intervals of time, it is more accurate to take referring expressions such as Scarlet O'Hara and Mr Spock to refer to entities which exist in the way real world individuals exist in w_0; that is to assume that these fictional individuals are made up of temporal stages which occupy a series of temporal counterparts of some W.

For fictional Ws which are set in counterparts of the real world, whether physically possible or not, we can distinguish those which are set in counterpart pasts, such as *Gone with the Wind* from those which are set in counterpart futures such as *Star Trek*. Figure 6.1 shows a plausible diagram of relative time-spans in three W's.

Figure 6.1

Relative time spans for three W's



It might be that syntactic signposts reflect the position of the time-span of the portion of some W which is described in a fiction in relation to our own w_0. If this were the case then we would expect to find past tense associated with *Gone with the Wind* and future modals associated with *Star Trek*.

There is another possible interpretation of tense signposts in sentences about fictional worlds:. Since a familiar world is one which a hearer has accessed (and assigned truth-values in) *previously*, it might be that the syntactic temporal signposts reflect the direction in time from the current w_0 to the temporal counterpart of w_0 from which the familiar world was first accessed. If this were the case then we would expect both *Star Trek* and *Gone with the Wind* to be associated with past tense.

The third possibility is that temporal signposts in sentences whose referents are located in other Ws, point from one w in *that* W to another. That is, they might require the hearer of some sentence to shift to a w_0 in the fictional W, in the same way as is necessary to allow them to interpret deictic expressions in fictions. If this were the case then we would expect both past tense and future modals associated with *Star Trek* and *Gone with the Wind*.

We do not intend to discuss the question of temporal reference in fictional worlds in great detail. However, let us consider brief example sentences which can be interpreted in both types of fictional world with syntactic signposts into the past and the future and with present tense verb forms. We will also consider a range of events and states. The various combinations of past or future fictions with temporal reference to events, and states of affairs in relevant W's and ws are shown in Tables 6.1 and 6.2.

Table 6.1

Past Fictions, temporal reference and states of affairs in possible worlds or sub-sets of possible worlds.

	events	**states of affairs in W**	**states of affairs in sub-set of W**
Past signpost	Scarlet O'Hara made a dress from curtains	Scarlet O'Hara was a Southern Belle	Scarlet O'Hara was married to Charles Kennedy
present signpost	Scarlet O'Hara drops a cup	Scarlet O'Hara is a Southern Belle	Scarlet O'Hara is married to Charles Kennedy
future signpost	Scarlet O'Hara will slap Rhett	Scarlet O'Hara will be a Southern Belle	Scarlet O'Hara will marry Charles Kennedy

Table 6.2

Future Fictions, temporal reference and states of affairs in possible worlds or sub-sets of possible worlds.

	events	**states of affairs in W**	**states of affairs in sub-set of W**
Past signpost	Mr Chekhov killed a Klingon	Mr Chekhov was the helmsman of the Enterprise	Captain Kirk was held prisoner on a frozen moon
present signpost	Mr Spock is re-born	Mr Chekhov is the helmsman of the Enterprise	Captain Kirk is not the commander of the Enterprise
future signpost	Mr Chekhov will kill a Klingon	Mr Chekhov will be the helmsman of the Enterprise	Captain Kirk will be held prisoner on a frozen moon.

The main observation to be made with respect to the sentences in tables 6.1 and 6.2 is that the pattern of acceptability of sentences with past and future signposts is the same whether the fictional W lies in the counterpart past or the counterpart future of w0. No matter whether the sentence describes an event, a state which holds throughout the entire time-interval of W described in the fiction, or a state which holds in some sub-set of W, signposts into the past are acceptable and signposts into the future less so. We conclude that w0 generally continues to be associated with the world the hearer of these sentences exists in and that the past syntactic signposts point the hearer into a world familiar to them from *their own past*. The future syntactic signposts on the other hand require the hearer to shift the identity of w0 of the W they exist in and into a w0' in the fictional W.

The use of the present tense with references to fictional entities appears to be consistent with references to entities in fictional Ws (i.e. in sentences which predicate of entities properties

which hold of them in all ws in the W in which they exist), and with references to entities in particular ws (i.e. in sentences which describe events which take place in one w of W[7]). In contrast, sentences which describe states of affairs whose duration is somewhere in between are somewhat strange with a verb in the present tense when the sentence is presented out of context. We conclude that to access a sub-set of W indicated with a present tense verb form it is necessary to shift the identity of w_0, to some w_0' which lies within the time-span of the fictional W which is described by the predicate.

We have now discussed the processes involved in accessing the current world, familiar worlds and various types of sign-posted counterparts. However, there are sentences whose referents cannot be located in any of these types of reference worlds. For example:

6.34. The King of France is bald.

The location of the referent of the subject NP of 6.34 is an example of the final type of world we will consider.

6.2.3.3. Last-resort Counterparts.

If no referent is can be located for some referring expression in w_0, or any familiar world, or by following a syntactic signpost, we propose that hearers interpret that referring expression as a reference to an entity in what we will call a *last-resort counterpart*. We will call the process by which these worlds are arrived at *step-by-step shifts* to distinguish them from the short-cuts and sign-posted shifts. Sentences 6.35-6.37 contain examples of referring expressions which must be interpreted in this way in the current context.

6.35. Camille Paglia's daughter has read *The Brothers Karamazov*.

6.36. Chelsea Clinton's sister painted her toe-nails magenta.

6.37. The Pope's wife celebrated Hanukkah quietly at home.

---

[7]that is in a sub-set of ws small enough to be indivisible or small enough for sub-dividing them to be unnecessary)

We assume that the reference worlds chosen in the interpretation of such sentences are minimal counterparts[8]; the reference worlds in 6.35-6.37 being minimal counterparts of the current w_0.

Again, we propose that shifts into last-resort counterparts of familiar worlds other than the real world are undertaken in precisely the same kind of circumstances. For example, last-resort counterparts are chosen as the existential locations of the referents 6.38 and 6.39, since no referents can be located any other way.

6.38. Scarlet O'Hara's fourth husband is a Yankee carpet-bagger

6.39. Sherlock Holmes' sister is kind but stupid.

We can observe that all three sentences in 6.35-6.37 and the pair of sentences 6.38 and 6.39 are perceived as truth-valueless in the current context. We claim that truth-valuelessness is a general feature of sentences containing only one referring expression, if that expression is interpreted by locating its referent in a last-resort counterpart world.

Clearly, we must defend this proposal. It might seem that it is contradictory to claim on the one hand that a particular world, used as a last-resort counterpart, is a very close counterpart of w_0, differing by perhaps only one property, but on the other hand that we cannot assign a truth-value to so much as one statement about that world. However, if we try to make a statement about one of the properties that a minimal counterpart of w_0 shares with w_0 then the utterance is interpreted as being about some entity, not in the counterpart, but in w_0 itself. For example, it is the case in a minimal counterpart where Chelsea Clinton's sister exists that Washington is the capital of the USA, but sentence 6.40 is not perceived as a true statement about an entity in that world but about an entity in w_0.

6.40. Washington is the capital of the USA.

---

[8]some implications for this assumption will be set out in section 6.4 in the discussion of cross-world relations

It might seem that a sentence such as 6.41. is a counter-example to this claim.

6.41. In the world in which Chelsea Clinton's sister exists, Washington is the capital of the USA.

However, this is not the case since the speaker of sentence 6.41 explicitly imposes the choice of reference world on the hearer, and the normal rules for choosing a reference world could not apply no matter what they were. The interpretation of 6.41 is in this way similar to the interpretation of sentences in story-telling discourse situations which do not conform to the general rules for interpretation, since every sentence is perceived as true[9].

It is interesting to observe that indicative conditionals which do not contain syntactic signposts can be interpreted by locating the referents of their referring expressions in last resort counterparts. For example:

6.42. If Camille Paglia's daughter is a Democrat, then Chelsea Clinton's brother is a Republican.

The important question to be addressed in an analysis of the interpretation of 6.42 is whether the referents of **Camille Paglia's daughter** and **Chelsea Clinton's brother** are located in the *same* counterpart world. Before we turn to a discussion of this issue and the consequent problems of cross-world relations, we will set out in full the order and application of the rules for locating referents for some simple cases.

---

[9]It is interesting that the suspension of the usual rules does not last indefinitely even in the context of story- telling. It would be extremely bizarre to react to an utterance of a. with an utterance of b..

a. Once upon a time there was a King. The King was getting rather thin on top.
b. No he wasn't.

However, it is not in the least strange that some reader might reject, for example, the actions predicated of Inspector Morse in the latest Colin Dexter novel, on the grounds that they are *out of character*. It appears that once a world is firmly established we begin to evaluate statements about it in a different way from that which we use when it is initially introduced.

6.3. Rules for Locating Referents: A Summary of Simple Cases.

We claim that the interpretation of a referring expression can be modelled as a process of locating a referent in some possible world. We also claim that the perceived truth-value for any sentence can be predicted on the basis of the choice of reference world and the level of epistemic access that the person assigning the truth-value has to that world. We propose that there are two constraints which apply to the location of referents:

C1. Begin the search in w_0.

Since discourses take place *in* worlds, w_0 at the outset of every discourse is the world the discourse participants exist in[10].

C2. Do not continue to search once a referent has been located and all syntactic sign-posts have been followed.

This constraint can be understood as an instruction to a hearer to locate a referent in w_0 or in a familiar world, or in a *minimal* counterpart of w_0 or a familiar world; i.e. that the reference world is chosen on the basis that the required entity exists there, whether or not the property which is predicated of that entity holds in the world or not. It is this constraint which ensures that every sentence is not interpreted as trivially true; the operation of this constraint is what allows us to predict that sentence 6.43, for example, is perceived as truth-valueless.

6.43. The King of France is bald.

The difference between story-telling discourse situations and normal contexts of interpretation can be defined as the suspension in story-telling discourses of both of these constraints.

Within these constraints, we propose that the location of referents occurs according to three

---

[10] We alluded to the possibility of changes in the identity of w_0 during the discussion of syntactic signposts in fictional sentences in section 6.3.2.2. and will return to this question and propose rules governing changes in the identity in section 6.4.

ordered rules.

R1. Use short-cuts.

This rule instructs a hearer to use the short-cuts associated with epistemic links to access established familiar worlds as locations for referents.

R2. Follow all signposts.

This rule instructs a hearer to access minimal counterparts if a syntactic signpost points into a modal or counterfactual world and to access temporal counterparts which lie in the direction indicated by temporal signposts.

R3. Locate a referent for every referring expression

This rule instructs a hearer to continue to apply the rules in order according to the first constraint until a referent is found.

We do not claim that these three rules all operate in every discourse situation. It is the suspension of all of these rules which characterises the interpretation of sentences given as evidence in court.

Let us examine the application of these rules to a group of sentences interpreted by a reader *x* in the current w_0 in the current context.

6.44. Noel Edmonds is surprisingly popular.

According to C1, *x* begins the search in the current world. A referent is available there and C2 ensures that the search therefore stops.

6.45. Inspector Morse is lamentably irascible.

Again C1 stipulates that the search begins in the current world. No referent is available there, and so *x* searches among his or her set of familiar worlds and locates a referent in a familiar world we can call w_{im} according to R2. C2 then ensures that the location process stops.

6.46. Inspector Morse might feel better if he drank less.

In the interpretation of this sentence, the referent of the name is located as for the referent in sentence 6.45. However, according to R2, the modal **might** acts a signpost into a minimal counterpart of w_{im}, and *x* locates the referent in w_{im}' which is an exact counterpart of w_{im} except that in w_{im}', the counterpart of Inspector Morse, Inspector Morse', drinks less than Inspector Morse in w_{im}.

6.47. Noel Edmonds might look better with a heavier beard.

The interpretation of the name in this sentence begins in w_0 according to C1, where a referent is located. The modal **might** then takes *x* out of w_0 into a minimal counterpart, w_0', in which the counterpart of Noel Edmonds, Noel Edmonds', has a heavier beard than Noel Edmonds in w_0.

6.48. Chelsea Clinton's sister might look good with a heavy beard.

Searching in w_0 according to C1 yields no referent for the definite **Chelsea Clinton's sister** and *x* cannot locate a referent in a familiar world by applying R1. However, by applying R3, which is triggered by the modal, *x* is taken to a minimal counterpart of w_0, w_0', in which an individual exists who is the sister of the counterpart in w_0' of Chelsea Clinton in w_0, and who has a heavy beard.

6.49. Chelsea Clinton's sister is lamentably popular.

The search for the referent of **Chelsea Clinton's sister** in 6.49, begins in w_0 according to C1.

No referent can be located by applying R1, since the definite does not pick out an entity in any of *x*'s familiar worlds, or R2, since there are no temporal or modal syntactic signposts into counterparts. R3, however, ensures that a referent is located, and the only possible location is among the set of last-resort counterparts, via a step-by-step shift. C1 ensures that the counterpart chosen is a minimal counterpart of w_0.

Having summarised the application of the constraints and rules of the location theory to some simple examples we will now consider the application of the theory to examples involving cross-world relations.

6.4. Cross-world relations.

The problems surrounding the location of referents in possible worlds become more interesting when we consider sentences containing more than one referring expression, for example,

6.50. A Dutchwoman danced with the King of France.
6.51. The King of France danced with a Dutchwoman.

We reviewed the arguments concerning the constraints on cross-world relations put forward by Fodor (1979) in section 2.3 of chapter 2. She addresses the questions in this area at some length, making some interesting proposals, but cannot account for the pair of sentences above using her set of competing pragmatic principles.

The question we must address here, is whether, with respect to sentence 6.51, having located a referent for the **King of France** in a last-resort minimal counterpart of w_0, w_0', we also try to locate the referent of **a Dutchwoman** in w_0', or whether we try to interpret **a Dutchwoman** as a reference to an entity in w_0. Clearly, if we attempt to locate a referent for the indefinite in w_0 we will succeed, since there are Dutchwomen in w_0. We can also be sure that if we attempt to locate a referent for **a Dutchwoman** in w_0', we will be able to find one there, since w_0' is a *minimal* counterpart of w_0. Given that C2 will ensure that the search ends whenever

a referent is located, the location of the referent for the indefinite is determined by where we look first.

Although we have no very reliable *direct* intuitions about the location of referents we can deduce the location of the referent of **a Dutchwoman** by consulting our intuitions about the resulting truth-value of the sentence. In this case, if we locate a Dutchwoman in w_0', then it is to be predicted that we will be unable to assign a truth-value to the sentence since it might be either true or false that the King of France in w_0' danced with a member of the set of Dutchwomen in w_0'. If, on the other hand, we locate a referent for **a Dutchwoman** in w_0 we predict that the sentence will be perceived to be false, since **dance with** expresses what Fodor called a same-world relation.

To decide which world we look in first, having shifted to w_0' to locate the referent of the subject NP, we must decide whether this world *becomes the current world* for the duration of the interpretation of the utterance (or possibly some larger chunk of discourse) or whether we "reset" at the *original* current world, the one in which the discourse takes place, to carry out the location procedure for the second NP.

We can suggest that during the interpretation of a sentence the identity of w_0 shifts. To reflect this we can redefine "the current world" as "the last world accessed". The interpretation procedures for 6.50 and 6.51 would be carried out in the following way.

Sentence 6.50.

SUBJECT NP	1.	Locate a referent in w_0 in accordance with C1
OBJECT NP	1.	Begin the search in w_0 according to C1.
	2.	Attempt to apply R2.
	3.	Locate the King of France in a last-resort minimum counterpart of w_0, w_0' by applying R3.
PREDICATE		Assign a truth-value of *false* using knowledge of the lexical selectional restrictions of **dance with** and the relative locations of the referents.

Sentence 6.51.

SUBJECT NP	1.	Begin the search in w_0 according to C1.
	2.	Search the set of familiar worlds in accordance with R1.
	3.	Attempt to apply R2.
	4.	Locate the King of France in a last-resort minimum counterpart of w_0, w_0' by applying R3.
OBJECT NP	1.	Locate a Dutchwoman in w_0' according to C1.
PREDICATE	Attempt to assign a truth-value by consulting knowledge of the properties of w_0'	

A central claim of the current thesis has been that the apparently exceptional status of the real world, in the interpretation of sentences, can be explained as a result of the fact that our discourses take place in this world and therefore the real world is often w_0. We will now consider a pair of sentences involving cross-world relations whose referring expressions do

not pick out entities in the real world. We will show that the interpretation of these sentences conforms to the same rules as the interpretation of sentences 6.50 and 6.51 above.

6.52. A Klingon had a dream about a Baker Street Irregular.

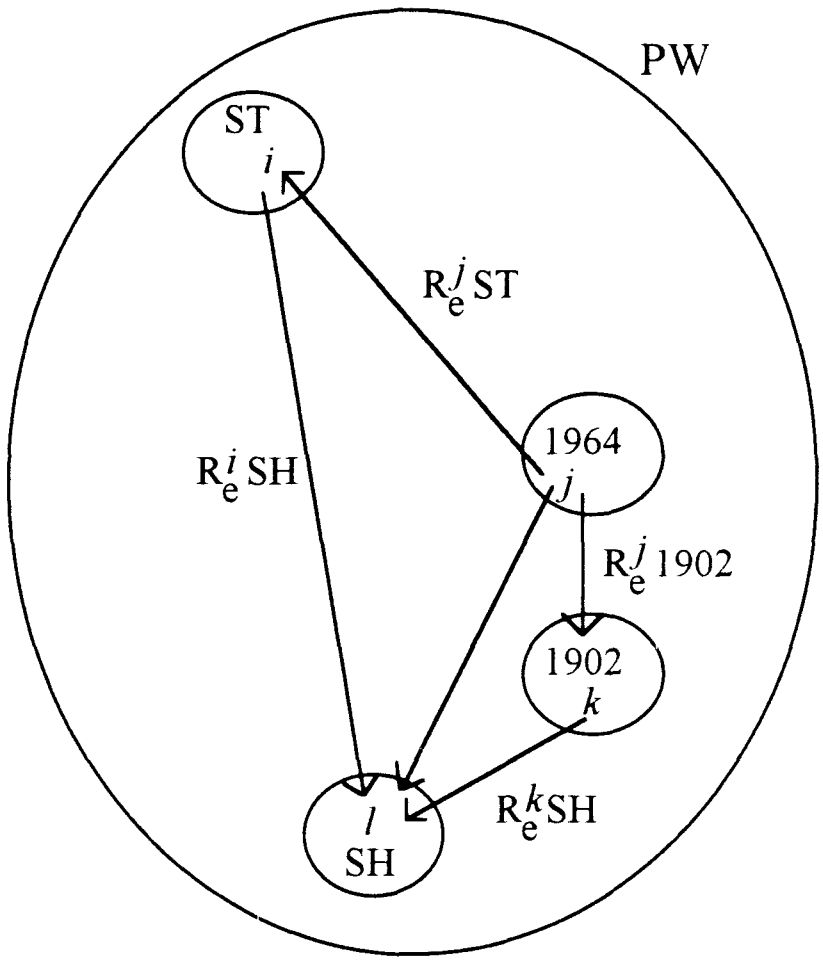
6.53. A Baker Street Irregular had a dream about a Klingon.

Recall that on Fodor's account based on opaqueness, two analyses are possible, neither of which is correct. If the expressions **A Klingon** and **a Baker Street Irregular** are taken to refer to non-existent entities (since their referents do not exist in the real world) then it is predicted that 6.52 and 6.53 will be truth-valueless, since failed referring expressions cannot occur as the subjects of asymmetric predicates. If, on the other hand, **A Klingon** and **a Baker Street Irregular** are assumed to pick out existent entities in fictional worlds then it is predicted that 6.52 and 6.53 are truth-valueless as a result of underspecification of fictional worlds. The truth-values suggested by intuition, that 6.52 is either true or false, but 6.53 is false, could only be produced if it were argued that Klingons exist but Baker Street Irregulars do not. Our account on the other hand correctly predicts the truth-values perceived. To show this we will examine the interpretation processes for each in detail

Consider the partial diagram of epistemically linked worlds in Figure. 6.2

Figure 6.2

Possible Epistemic links between the worlds of Star Trek and Sherlock Holmes .



ST = the world invoked by *Star Trek*

SH = the world invoked by the fiction of Arthur Conan Doyle

i = a member of the set of Klingons

j = Gene Roddenberry

k = Arthur Conan Doyle

l = a member of the set of Baker Street Irregulars.

The assumption appears to be well-motivated that to dream about an entity, it is necessary to have an epistemic link into the world where that entity exists.

In the interpretation of 6.52 in 1996, the application of C1 yields no referent for **A Klingon** in w_0. R1, however can be used to locate a set of Klingons in a familiar world w_{st} and the

referent of **a Klingon** can be interpreted as a member of that set. The next stage in the interpretation procedure is to locate a referent for **a Baker Street Irregular.** Whether we take the current world to be w_0 or w_{st} C1 cannot yield a referent. We then apply R1. If we try to apply R1 to *i*'s familiar worlds, we cannot arrive at an answer. It may be that some world which contains a referent is one of *i*'s familiar worlds, but we do not know, since we do not know which worlds are connected to *i* via epistemic links. On this interpretation we do not know whether the sentence is true or false. If we apply R1 to our own familiar worlds we locate the set of Baker Street Irregulars in w_{sh}. We locate the referent of **a Baker Street Irregular** in this set. We know that there *might* be an epistemic link from *i* in w_{st} to w_{sh} since we know that w_{sh} is epistemically accessible from w_{st}. We know *this* because we know the initial epistemic link $R_e{}^kSH$ is from 1902, a counterpart of which is epistemically available to individuals in w_{st} since it is an *earlier* temporal counterpart. Although we know that an epistemic link is possible we do not know whether such an epistemic link actually exists. Therefore we do not know whether the sentence is true or false

In contrast, in the interpretation of 6.53 the application of R1, after the failure to locate a referent in w_0 in accordance with C1, yields a referent for **a Baker Street Irregular** in w_{sh} R1 yields no referent for **a Klingon** in either w_0 or w_{sh}. We then attempt to apply R2 to locate a referent for **a Klingon** in a world epistemically linked to l in w_{sh}. In this case, we know that w_{st} cannot be epistemically linked to l in w_{sh}, since the *initial* epistemic link into w_{st} is out of 1964, and counterparts of 1964 are not epistemically available to individuals in w_{sh} since they are *future* temporal counterparts of this world. We can conclude, that sentence 6.53 is false, since it is not possible for an individual to have a dream about a second individual which is not epistemically available to him or her.[11]

---

[11]On the reading where we search the familiar worlds of other-worldly entities, it is *just* possible to get the reading that in the world where Sherlock Holmes and the Baker Street Irregulars exists someone creates an epistemic link to the world of Star Trek (and by a massive coincidence Gene Roddenberry follows the same path in 1964). This "possibility" is indigestible enough to suggest that the other alternative is the correct one; i.e. that we search our own familiar worlds.

6.5. Summary.

In this chapter, we showed how the relation of relative proximity between worlds which results from the structure we proposed in chapter 5 translates into an ordering of worlds in terms of relative accessibility. We decided to establish the world in which a discourse takes place as the nearest and hence most accessible potential reference world for the referring expressions in the utterances of that discourse. From this basis, we proposed that the set of epistemic links which connects an individual in one world to other worlds should be seen as a system of short-cuts and this allowed us to define the familiar worlds for some individual as the next nearest and hence the next most accessible set of locations for referents. We divided the remainder of the set of worlds, which were defined as counterparts, in terms of their conditions of use as locations for referents; distinguishing between counterparts which are employed in accordance with modal or temporal syntactic signposts, and counterparts which are employed as a last resort in the attempt to locate referents.

Having discussed the possible methods for locating referents, we then set out to predict how referents are located in actual discourse situations, by setting out an ordered set of constraints and rules which govern the choice of reference world during interpretation. We showed that the rules correctly predict the existential locations for a range of referring expressions in simple sentences and that the resulting patterns of truth-value assignments and truth-valuelessness accord with our intuitions. We then moved on to look in detail at more interesting cases involving pairs of expressions which could be analysed as having referents in different worlds. We found that certain sentences which had been exceptional under previous theories could be accounted for by the rules for locating referents we proposed.

Chapter 7.
Problems and Implications.

7.1. Introduction.

In this final chapter we will examine the potential problems encountered by adopting the location theory of reference presented in this thesis. These problems are of various types and degrees of seriousness.

First we will examine two individual types of NPs, deictic pronouns and names, which do not appear to follow the general rules for interpretation which we have proposed to account for the entire set of NP types. It will be shown in sections 7.2.1. and 7.2.2. that these apparent counter-examples can be accounted for by appealing to their specific locations on the EF scale.

Second, we will turn our attention to a discussion of a potentially undesirable consequence of adopting the view that nominal expressions in so-called opaque contexts can be analysed as picking out referents rather than intensions: this stance leaves us in need of an explanation for the failure of the Law of Substitution in these contexts. A detailed solution to this problem is outside the scope of the present work but we make an initial suggestion for an analysis which appears promising in section 7.3.

Lastly, we take close and critical look at the structures we proposed to define sets of temporal counterparts[1]. We acknowledge that the theory entails some quite radical departures from the existing, accepted ideas about temporal reference in other theories (Cann, 1993, Kamp and Reyle, 1993). Since the structure of the set of possible worlds which we have presented was developed with the central aim of accounting for nominals, it is likely that some far-reaching modifications might be necessary to enable the theory accurately to handle a range of predicate types.

---

[1]see Chapter 5, section 5.2.

7.2. Problem Cases.

Fodor identified the following two related problems: (1) that deictic pronouns appear to trigger obligatory real world reference (e.g. in sentence 7.1.) and; (2) that it is difficult to account for which *other* referring expressions behave like deictic pronouns with respect to world shifting constraints (e.g.. sentence 7.2.) and which allow free shifting between worlds (e.g.. sentence 7.3.).

7.1. My mother sat next to Hercule Poirot on a flight between Paris and London.

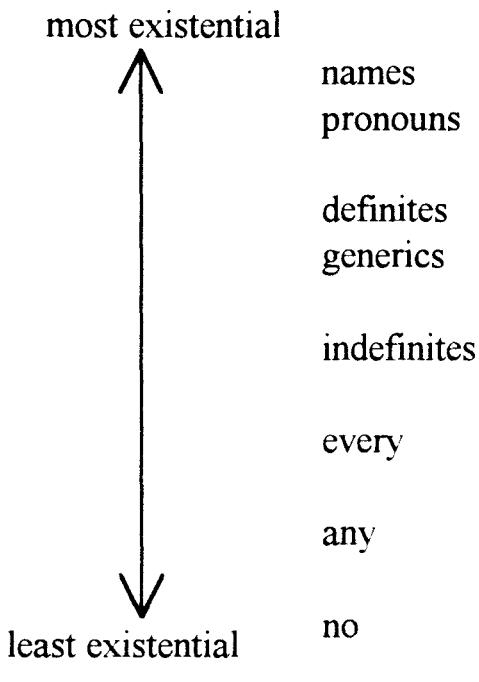
7.2. The Chippendales danced at the garden party in St Mary Mead.

7.3. A Chippendale chair was stolen during the garden party at St Mary Mead.

Having proposed that the set of determiners be ordered on a scale of existential strength, it would provide additional support for the usefulness of this scale, if we could show that it is possible to identify some point on the scale as the precise location of the boundary between entities whose referents are fixed in the real world and entities which can refer in other worlds. In the following two sections, we offer an explanation which *does* account for both of these problem areas in terms of the scale of EF. A partial representation of the scale we proposed is given again in Fig. 7.1:

Figure 7.1.

Scale of existential force for referring expressions



We proposed that the effect of the strong EF associated with expressions at the upper end of this scale is to trigger the analysis of such expressions as references to entities whose existence is pre-established. We defined such pre-established entities, with respect to some discourse participant, as entities whose existence and identity is established for that individual prior to the utterance of the current reference to them. We are now in a position to restate this requirement of pre-established existence in terms of possible worlds. We can say that expressions with the strongest existential force will be interpreted by a hearer as references to entities which are located in pre-established worlds; and this in turn can be seen as a consequence of the general requirement to locate referents in the nearest available world. The nearest world is either w_0, the world in which the discourse takes place, or a member of the set of familiar worlds with which the hearer has epistemic links. What remains to be decided is how far down the scale this effect reaches.

7.2.1. Pronouns

The first place on the scale of existential strength, occupied by pronouns, appears to constrain world-shifting extremely tightly. We can show that pronouns are always interpreted in one of two ways neither of which involves any shifting of reference world . For example:

7.4 The King of France cut his toe-nails.

If 7.4 is a discourse initial utterance, then the only possible reading is one on which **his** is coindexed with **the King of France,** i.e. a reading where the referent of the pronoun exists in the world last accessed. The pairs of sentences in 7.5 and 7.6 indicate that no such consideration as Fodor's Real World Principle[2] either overrides or applies with equal weight as the rule for interpreting the referents of pronouns in the last world to have been accessed.

7.5a. Mrs Thatcher ironed Dennis's$_i$ trousers. The King of France$_j$ cut his$_j$ toe-nails.

7.5b. Mrs Thatcher ironed Dennis's$_i$ trousers. The King of France$_j$ cut his$_i$ toe-nails.

7.6a. The Queen of France ironed the King's$_i$ trousers. Dennis Thatcher$_j$ cut his$_j$ toe-nails.

7.6b. The Queen of France ironed the King's$_i$ trousers. Dennis Thatcher$_j$ cut his$_i$ toe-nails.

If a RWP *did* operate and took precedence over the rule we propose then only the co-indexings represented in 7.5b and 7.6a should be possible. If an equally weighted RWP operated then both variants in 7.5 but only 7.6a should be possible. If neither constraint applied then all four co-indexings should be possible. The fact that only 7.5a and 7.6a are available readings indicates that the rule for locating the referents of pronouns constrains the choice of reference world to the last world accessed.

---

[2]see Chapter 2, section 2.3.4.

In cases where this type of interpretation is not available for a pronoun, the requirement still holds that referents are located in the nearest world. For deictic pronouns, the nearest world, in almost every case, is the world in which the speaker exists, w_0, and deictic first person pronouns will therefore almost always be interpreted as references to the speaker in the world in which the discourse takes place. For example, sentences such as 7.7 and 7.8 will, we claim, *always* be interpreted as false statements about entities in separate worlds from each other, no matter in what order the two expressions occur;

7.7a My mother went to school with the King of France.

7.7b The King of France went to school with my mother.

7.8a I went to school with Chelsea Clinton's sister.

7.8b Chelsea Clinton' sister went to school with me.

The referents of **me** or **my mother** in the b versions cannot be located in the last-resort counterparts accessed to interpret the **King of France** and **Chelsea Clinton's sister**, even though these worlds have just been accessed. The reason for this is that the referents must be located in the nearest world in which they exist; i.e. the world in which the discourse takes place, the current world. However, even when the nearest world in which the referent exists is not the current world, as is the case in sentence 7.9, the referent cannot be located in the last accessed counterpart and the sentence is perceived as false.

7.9. Sherlock Holmes' wife went to school with my great-grandmother.

The explanation for this constraint is that the referent of **Sherlock Holmes' wife** must be located in a last resort-counterpart, since none is available in any familiar world . There is no referent for **my great-grandmother** available in the current world. However, the nearest world where such a referent is available is a temporal counterpart of the current world which is among my set of familiar worlds. That this world is nearer than the world where we located Sherlock Holmes' wife is a result of viewing epistemic links as short-cuts. Even though the

world where Sherlock Holmes wife exists has been accessed it is not a familiar world since it has not been accessed *prior* to this utterance.

There is a further class of apparent exceptions which give additional strength to an analysis in terms of close, familiar worlds rather than a Real World Principle, since they provide more evidence that we can use deictic items in other worlds under circumstances which ensure that these other worlds are familiar according to the definition of familiarity given in section 5.4, Chapter 5. First, consider 7.10 which is a quotation by an actress discussing a role in a film; i.e. a fictional world in which she exists, and in which the values of properties holding of her are explicitly specified.

7.10 She's$_i$ an interesting character - that is to say I$_i$ got killed after ten minutes but I$_{sp}$ spent seven of them in Tom Cruise's arms[3].

Since the speaker cannot have been killed, the first **I** must be coindexed with **she**, but since Tom Cruise is not a inhabitant of the fictional world where the character exists, the second **I** must refer deictically to the speaker. In contrast 7.11, in which a familiar world is not available, is simply false, since the first person deictic pronoun cannot, according to the restriction we are proposing, refer to an entity in an unsignposted counterpart.

7.11. I had an interesting war - that is to say I got killed after ten months, but I spent seven of them in a beautiful spot in the Mediterranean.

We claim that if there is a potential choice, in interpreting an expression of weaker existential force, between a non-existential reading which does not contravene the restrictions on familiar world reference of deictic pronouns, and an existential reading which does contravene these restrictions, the non-existential reading is preferred. This contrasts with the general case where existential readings are preferred to non-existential readings. For example, the QNP **every child of mine** in sentence 7.12, in a context in which the speaker has no children, can either be interpreted existentially in a counterpart world in which the speaker has children, or

[3]The Times 9/12/95

non-existentially in w_0. We take the acceptability of 7.12a and the unacceptability of 7.12b, as potential continuations of 7.12, to indicate that the non-existential, current world reading is preferred.

7.12 Every child of mine can play the violin.

7.12a. although my ears have suffered less than Ginnie's have since her two started learning.

7.12b. It's rather painful at the moment, but they are getting better.

In contrast, in 7.13 the preferred interpretation for **everything in sight** is existential, as indicated by the fact that 7.13b, rather than 7.13a, is the preferred continuation of the two.

7.13 I drank everything in sight.

7.13a. - thankfully for my liver the place is not licensed.

7.13b. - which might have tipped me over the Government's new safe daily level.

We claim that temporal deictic items also have an effect on the acceptability of references to entities in other worlds. If we take tense to be the least deictic type of temporal construction and temporal adverbials, which place reference times very close to utterance times, as the most deictic type, we can see that there is a corresponding cline of likelihood that we will interpret referring expressions as references to entities in other worlds. For example, we claim that sentences 7.14 - 7.17 are increasingly likely to be interpreted as false, indicating that the object NPs are increasingly likely to be interpreted as references to entities in the real world.

7.14. The King of France danced with a Dutchwoman.

7.15. The King of France danced with a Dutchwoman last year.

7.16. The King of France danced with a Dutchwoman last night.

7.17. The King of France danced with a Dutchwoman an hour ago.

We conclude that at the strongest existential point on the scale, deictic pronouns (and other deictic elements) must refer in the nearest world to the speaker, which is almost always the

world in which the discourse is taking place and otherwise is a familiar world or a signposted counterpart.

7.2.2. Names

After a detailed discussion in section 3.1 of Chapter 3 we concluded that all occurrences of names should be analysed as having the strength of EF which calls for their interpretation as references to pre-established entities. We suggested that the utterance of any of the versions of 7.18 (by a speaker to an unfamiliar hearer), which might be thought of as a counterexample to this claim, is only possible given the assumption that individuals typically (or at least stereotypically) have partners, children, friends and a network of acquaintances, and that each of these has a name.

7.18 I didn't sleep last night - because Norma is away and I was cold.
because Murdo is teething.
because Petra kept phoning from Japan.

With regard to the freedom of choice for the reference world of a name, we claim that they are even more tightly restricted than pronouns. Although we agree with Kripke's analysis of names as rigid designators, which therefore pick out the same individuals (or counterparts of the same individuals) in all worlds, (Kripke, 1972) we argue that names are always *interpreted* as referring in the nearest world, which will always be a familiar world. There is no incoherence in this view since we have already accepted that although counterparts of entities exist in an infinite number of possible worlds, referring expressions are typically interpreted as references to a small number of possible counterparts in a very few of these worlds.

We predict that a combination of a pair of names which pick out entities in different familiar worlds with a predicate which requires physical co-existence will always result in a truth-value of *false*; as shown in the examples in 7.19 - 7.22.

7.19. Barbara Cartland spat at Lieutenant Uhuru.

7.20. Lieutenant Uhuru spat at Barbara Cartland.

7.21. Lieutenant Uhuru spat at Scarlet O'Hara.

7.22. Scarlet O'Hara spat at Lieutenant Uhuru.

Asymmetric predicates combined with named entities in different worlds will function in the same way as the predicates discussed in section 6.4 in Chapter 6 in such sentences as 7.23.

7.23. A Klingon dreamed about a Hobbit.

That is, sentence 7.24 will be perceived as truth-valueless, which indicates that it is potentially true, while sentence 7.25 will be perceived as false.

7.24. Mr Chekhov dreamed about Sherlock Holmes.

7.25. Sherlock Holmes dreamed about Mr Chekhov.

Even names which are explicitly introduced to identify individuals in sign-posted counterparts can never force names which *do* refer to entities in familiar worlds to shift worlds. This is indicated by the falsity of 7.27 and 7.28 as continuations of 7.26 (which is potentially true).

7.26. Chelsea Clinton's brother might be called **Hampstead Clinton**.

7.27. and Hampstead Clinton is Chelsea Clinton's brother.

7.28. and Hampstead Clinton danced with Sue-Ellen Ewing.

We predict on the basis of this observation that the only possible truth-value for sentence 7.29 is *false* since **Paris** must be taken as a reference to an entity in the current w_0.

7.29. The King of France lives in Paris.

Recall that Fodor uses this sentence as an example of *permissible* world-shifts of entities. We do not agree with her evaluation, but concede that a marginal reading of this sentence as potentially true is possible, if it is interpreted as a (slightly weird) generic with the expression **the King of France** being interpreted as a generic reference to the entire set of Kings of France, or a stereotypical representative of all Kings of France. On this reading it is irrelevant that the current temporal counterpart of our W does not contain one of these Kings of France to live in our counterpart of Paris. Note that sentence 7.29 is much more acceptable as a true statement than 7.30.

7.30 The King of France holidays in Honolulu.

It is not plausible to claim that this difference in our intuitions about sentences 7.29 and 7.30 exists because the referent of **Honolulu** is generally less likely to be located in a counterpart world than the referent of **Paris.** We prefer to argue that 7.30 is more likely to be interpreted as potentially true than 7.29 because it is a more reliable stereotype that Kings of France live in Paris than that Kings of France holiday in Honolulu[4]. This is borne out by a comparison of 7.31 (with a forced generic interpretation of the subject NP) and 7.32 (in which the subject NP can only be given a specific reading).

7.31. The King of France typically lives in Paris.

7.32. The King of France lives in Paris now that his wife is dead.

The intuition is that 7.31 is much more acceptable as a true statement than 7.32.

To summarise, a dividing line must be drawn between expressions which are existentially strong enough to force the location of their referents in the nearest (and hence familiar) worlds and expressions which follow the general rules for world shifting laid out in the location theory. Our proposal is that this line should be drawn at a point on the scale below names and

---

[4]McCawley noted the availability of this kind of generic reading for sentences such as 6.18. see Chapter 6 section 6.2.3.1.

6.18. If London is in England, the Queen of England lives in London

pronouns and above the remainder of the expressions which we analysed as being associated with pre-established EF; that is the entire class of definites and generics.

To show that this is the correct analysis, let us consider those expressions, such as **the Dalai Lama,** which can be interpreted either as names, i.e. as rigid designators, or as definite descriptions. For example, **the Dalai Lama** functions as a name which picks out an individual (the 14th incarnation) in the current w_0 in 7.33 and picks out counterparts of that individual in signposted counterpart worlds in 7.34 and 7.35.

7.34. The Dalai Lama has a kind face.

7.35. The Dalai Lama might never have become a monk.

7.36. If the Dalai Lama had been a girl, he would have been barred from religious life.

In contrast, **the Dalai Lama** in 7.37 and 7.38, functions as a definite description and picks out a generic or stereotypical referent.

7.37. The Dalai Lama has not such had a kind face every time.

7.38. If the Dalai Lama was not a monk, Tibetan Buddhism might be a very different kind of religion.

Because of this ambiguity, we claim that if 7.39 is to be interpreted as a potentially true sentence, i.e. for **the Dalai Lama** to be interpreted as referring to an entity in the same counterpart world where we locate the referent of **the King of France**, **the Dalai Lama** cannot refer to a counterpart of the current 14th incarnation who exists in w_0. If **the Dalai Lama** refers to the 14th incarnation then sentence 7.39 is interpreted as false.

7.39. The King of France shook hands with the Dalai Lama.

Likewise, if sentence 7.40 is to be interpreted as potentially true, then **the President's daughter** does not refer to a counterpart of Chelsea Clinton.

7.40. The King of France danced with the President's daughter.

We conclude that in the absence of any context which suspends the normal rules for interpreting referring expressions[5], or any syntactic signposts into counterparts, these expressions when they are interpreted as the names of individuals in w_0, cannot be interpreted by locating their referents in last-resort counterparts. On the other hand when they are interpreted as referring expressions their referents *can* be located in the last-resort counterparts, accessed previously, if suitable referents are available there.

7.3. More Opaqueness: The Intensional/Extensional Distinction and the Failure of Leibniz's Law.

Besides the cancelling of presuppositional effects which we discussed in section 4.3.1 of Chapter 4, there is a second argument which is used to support the notion that referring expressions do not denote their extensions in opaque contexts and that is that in these contexts Leibniz's Law, the law of substitution, fails. This law states that if some expression is substituted for another semantically equivalent expression in a statement, the truth-value of that statement is unchanged. For example, if 7.41 is paraphrased as 7.42, Leibniz's Law dictates that it continues to be true in the current context.

7.41. Prince Philip charmed the ambassador.

7.42. The Duke of Edinburgh charmed the ambassador.

However, in opaque contexts, e.g. following verbs such as **believe**, **dream about**, **want**, and **think**, the law of substitution appears to fail. For example, it is possible to read 7.43 as false and 7.44 as true in the same context.

7.43 Camille wants to meet the Duchess of York .

---

[5]such as explicit story-telling like **Long ago and far away . . .**

7.44 Camille doesn't want to meet Fergie.

That is, it may be true that Camille, who does not know what individual **the Duchess of York** picks out, either mistakenly thinks the Duchess of Kent_i is the Duchess of York and wants to meet *her*$_i$, or simply wants to meet the Duchess of York whoever she may be because, say, she once had a pony called The Duchess of York (i.e. a de dicto reading). Alternatively it may be the case that Camille saw the individual who *is* the Duchess of York and wants to meet that individual (i.e. the de re reading).

It may be true in exactly the same context, that Camille does not know what individual **Fergie** picks out, but thinks Fergie is a royal corgi and does not want to meet her because she is scared of dogs, or thinks the name **Fergie** sounds like the name of someone unpleasant and does not want to meet her whoever she may be (i.e. a de dicto reading). It can only be possible for Camille to have seen the individual we can identify as Fergie and not want to meet that individual (on a de re reading) if she believes that the individual she does not want to meet is a different individual from the individual she does want to meet as described by 7.43 on the de re reading.

One answer to the problem of Camilles's conflicting beliefs/desires is to analyse the definite or name on the de dicto reading as denoting its intension, regardless of the (contingent) identity of the extension, and to analyse the same expression on the de re reading as denoting its extension. This kind of analysis is clearly incompatible with our claim that the interpretation procedure for a referring expression such as a definite or a name is the procedure of locating its referent; i.e. its extension.

The puzzles thrown up by opaque contexts and the failure of Leibniz's Law are bound up with the fact that no matter how many ways there are of referring to an entity, it is always the same entity. The advantage of our theory is that we distinguish the *fact* that referring expressions must pick out their referents, and that statements must have truth-values, from the *possibility* that not all individuals can locate the referents of names and other referring expressions, or decide upon these truth-values. That is, we recognise the effect that epistemic limits can have

on acts of interpretation and truth-value assignment. For this reason, the fact that Camille might have restricted knowledge of, or mistaken beliefs about, the connection between either of the two expressions **Fergie** and **the Duchess of York** and an individual in the world, is separate from, and does not inhibit the ability of, the *hearer* of 7.43 or 7.44 to pick out the referent of either of these expressions during the interpretation of sentence 7.43 or 7.44. We recognise the difference between Camille having conflicting desires or beliefs and Camille (or anyone else) **knowing** that she has conflicting desires or beliefs.

There is a further benefit of adopting Leibniz's Law for a certain type of semantic theory; those theories on which all expressions without referents are taken to denote the same entity, the empty set (e.g. Montague Grammar, Dowty *et al.*, 1981; Cann, 1993). According to such theories, the two sentences 7.45 and 7.46 are equivalent (and false) if the indefinite descriptions in object position are interpreted extensionally.

7.45 Norman had a dream about a unicorn.

7.46 Norman had a dream about a phoenix.

If, on the other hand, the two indefinites in 7.45 and 7.46 are taken to denote their intensions then each sentence can be assigned a distinct truth-value. This is an important advantage of an intensional analysis over an extensional analysis **if** the extensions of these expressions are assumed to be the empty set.

From the point of view of location theory, however, nothing is gained by attempting to interpret such expressions as denoting their intensions. The problem of extensional equivalence simply does not arise if the extensional referents of these indefinites are taken to be entities (unicorns and phoenixes) which exist in non-real counterpart worlds and which are identified using the rules of the location theory. The assignment of truth-values to 7.45 and 7.46 depends on the current reader's ability to identify the referent of **Norman.** Assuming that Norman can be located in the current world, the assignment of either truth-value is possible (so long as the reader has the necessary epistemic access to information about Norman) since **dream about** describes a cross-world relation. We assume that, in fact, these sentences are

left truth-valueless by the current reader but this in itself indicates that they are potentially true. If this were not the case, the sentences would be assigned the truth-value false as is sentence 7.47.

7.47. Elizabeth I had a nightmare about Windows 95.

Clearly, further investigation is required into the wider implications of abandoning the intensional/extensional distinction, but we consider the current analysis encouraging enough to regard this a worthwhile exercise.

7.4. Potential Problems with Temporal Counterparts.

Finally, we must consider two types of potential problems that can be foreseen for a location theory of interpretation based on the view of the set of possible worlds set out in Chapter 5. These problem areas are the result of first, the split between the interpretation of referring expressions and the interpretation of sentences and second, the realistic view we take of the constraints placed on interpretation by the epistemic limits of the interpreter.

The model we have proposed for the interpretation of utterances of sentences is a procedural one with two distinct stages ordered one after the other: first, the location of referents and second, the fixing of truth-values. It is a central claim of the theory that the location of referents takes place prior to the evaluation of the truth or falsity of the properties being predicated of them in the sentences in which the references to them occur[6]. The problem this view causes is that nothing about the referring expressions in a sentence gives any clue to the hearer of that sentence whether the referents must be located in instantaneous worlds without temporal parts -single ws - or in worlds which persist through some interval of time - sub-sets of Ws containing strings of temporally related ws. For example, the interpretation of sentence

[6]Chapter 6, section 6.1.1.

7.48 requires the existence of a referent in only a single w, while the interpretation of 7.49 requires the existence of a referent in a set of ws over an extended interval of time.

7.48. Diana looked bashful.

7.49. Diana is gradually amassing power and public sympathy.

Since it is possible to interpret the name **Diana** as a reference to an example of the type of entity, a human individual, which perdures through time (Lewis 1986), there may be no insurmountable technical problems to prevent an interpretation procedure in which a referent is located first in one individual w, as a result of the interpretation of the nominal expression, and then the range of reference worlds is expanded during the interpretation of the utterance as whole. Very much more research is required before a coherent answer is likely to emerge. We would need to look in detail at the data which is handled by those theories such as Montague Grammar and DRT which do seriously address the problems of temporal reference, before we could evaluate the scale and seriousness of the problem for the current theory.

The final potential problem we will consider is the extent to which a hearer's limited epistemic access to counterpart worlds constrains his or her ability to assign truth-values. We count it as an empirical strength of the theory that it predicts that hearers are largely unable to assign truth-values to factual statements about entities located in last-resort counterparts and we have discussed truth-value assignment in relation to counterfactuals and modals at some length in section 6.2.3 of Chapter 6. Here we will consider briefly how limited access to temporal counterparts might constrain the assignment of truth-values to sentences containing verbs of any of the distinct types identified by Vendler: states and actions (activities, accomplishments and achievements) e.g. 7.50 - 7.53 (Vendler, 1967).

7.50.	Joan is an actress of distinction.	STATE.
7.51.	Joan puts on her mascara.	ACTIVITY.
7.52.	Joan won her case.	ACHIEVEMENT.
7.53.	Joan is writing a book.	ACCOMPLISHMENT.

We claim that the reference worlds for some utterance are chosen on the basis of their relative accessibility or closeness to the world in which the hearer of the utterance exists. We also claim that familiar worlds, those worlds with which hearers have epistemic links can be defined as more accessible than other counterparts[7]. We argued that for some world w_0 at time t earlier temporal counterparts of w_0, i.e. w_0' at time t-1 etc., can be counted among the familiar worlds of individuals who exist at w_0, but that later counterparts of w_0, i.e. w_0' at time t+1 etc., cannot be defined as familiar worlds of these individuals. In fact, we argued that the later temporal counterparts of any w_0 are indistinguishable from other general counterparts of w_0 from the point of view of the individuals at w_0.

An interpretation of 7.53, which contains an accomplishment verb, in a Montague Semantic framework, would state that the proposition expressed in 7.53 is true with respect to some index at time t if Joan begins writing a book at some time t-n and finishes writing a book at time t+n. In the terms of the location theory, however, this means of assigning a truth-value to sentence 7.53 is impossible, since a hearer of 7.53 cannot have the necessary epistemic access to a future temporal counterpart of w_0 at any time t+n which will allow him or her to establish the completion at that time of the accomplishment described by the verb. We must provide another means by which a hearer of 7.53 can assign the truth-value true (or false) to this sentence. This may not prove to be a serious problem; an initially promising approach is to argue that if a hearer can establish the beginning point of this accomplishment in an earlier temporal counterpart of w_0 and is unable to identify any intervening temporal counterpart between *this* point and w_0 he or she can conclude that the period during which the writing event takes place does include w_0 and hence the sentence can be interpreted as true. This is a radical departure from the view of the interpretation of this sentence taken by Montague Grammar and DRT and much additional research is necessary into the implications of limited epistemic access for the interpretation of different event types before a serious comparative evaluation of the approaches can be attempted.

---

[7]Chapter 5, section 5.4.

7.5. General Conclusion.

The problem we set out to address in this thesis was how to account for the pattern of perceived truth-values for a range of sentences which use a variety of types of nominal expressions to pick out various types of entity.

It was clear that although numerous semantic and pragmatic presuppositional theories purported to address precisely this area the problems involved were at the same time too closely bound up with truth and falsity to be handled by pragmatic theories in terms of appropriateness or felicity, yet too much influenced by factors external to the linguistic system to be accounted for in terms of inherent truth-values for sentences.

By developing a theory which takes account of the epistemic states of speakers and hearers in discourse situations, and views truth-value assignment as a pragmatic process, separate from any statement of truth-conditions, we hoped to predict perceived truth-valuelessness for sentences without tampering with the bi-valent logical account of their semantics.

To do this we reviewed the efficacy of existing approaches to the existential features of linguistic expressions which claim that existential import is a binary phenomenon. We proposed instead that nominal expressions should be placed on a scale of existential strength allowing us to express the distinction between definites and indefinites in terms of existential force and to reflect the intuition that some NP types, traditionally held to be non-existential, could have existential interpretations. In this way we were able to analyse a wider range of nominals and nominal uses as expressions for which referents must be located.

To provide a framework for the required range of types of existence associated with entities we set out to construct a model of existential locations using the notion of an infinite set of possible worlds. We employed a realistic view of the nature of possible worlds on which all worlds are seen as actually existing entities, all though other worlds are assumed to be temporally, spatially and causally unconnected to the world in which we exist. By adopting this stance, we are forced to countenance the existence of every possible world and the

consequent truth somewhere of every sentence. To constrain the use of possible worlds in interpretation in such a way as to reflect the actual truth-values which are assigned, we imposed a structure on the set of worlds which ordered worlds in terms of their accessibility; the organising principle of the structure being that the level of accessibility of some world to individuals in another is governed by the degree of proximity between the two worlds.

The final goal of the thesis was to show that the combination of the scalar view of existential force and the rich structure of relations between worlds allowed us to reflect perceived truth-values. We examined a variety of sentences and demonstrated that their interpretations could be explained in terms of the worlds selected as the locations for the referents of their constituent nominal expressions. Given the fairly complex view of the relations between worlds, we were able to state very simply the rules which appear to govern the choice of reference world during interpretation. We showed that these simple rules were able to account for the perceived truth-value assignments of a wider range of sentences than could be handled by previous theories.

Bibliography

Abney, S (1987) The English Noun Phrase in its Sentential Aspect. PhD Thesis, MIT, Massachusetts.

Allwood J, Andersson, L-G and Dahl, Ö (1977) *Logic in Linguistics*. Cambridge University Press, Cambridge.

Austin, J L (1962) *How to Do Things With Words*. Clarendon Press, Oxford.

Barwise, J and Cooper, R (1984) Generalised quantifiers and natural language. *Linguistics and Philosophy* 4: 159-219.

van Benthem, J (1986) *Essays in Logical Semantics*. Reidel, Dordrecht.

Bloom L, Lightbown and Hood (1978) Pronominal-Nominal variation in Child Language. In Bloom, L (ed.) *Readings in Language Development*. Wiley, New York: 231-238.

Bowerman, M (1973) *Early Syntactic Development*. Cambridge University Press, Cambridge.

Bradley, R and Schwarz, N (1979) *Possible Worlds*. Basil Blackwell, Oxford.

Brown, G and Yule, G (1983) *Discourse Analysis*. Cambridge University Press, Cambridge.

Burton-Roberts, N (1989) *Limits to Debate: A Revised Theory of Semantic Presupposition*. Cambridge University Press, Cambridge.

Cann, R (1993) *Introduction to Formal Semantics*. Cambridge University Press, Cambridge.

Carlson, G N (1977) *Reference to Kinds in English*. PhD Thesis, University of Massachusetts.

Carlson, G N (1989) English generic sentences. In, Chierchia, G, Partee, B H & Turner, R (eds.) *Properties, Types and Meaning: Volume 1 Foundation Issues*. Reidel, Dordrecht.

Carnap, R (1947) *Meaning and Necessity*. University of Chicago Press, Chicago.

Chierchia, G and McConnell-Ginet, S (1990) *Meaning and Grammar: an introduction to semantics*. MIT press, Cambridge, Massachusetts.

Crystal, D (1987) *The Cambridge Encyclopedia of Language*. Cambridge University Press, Cambridge

Dowty, D R, Wall, R E and Peters, S (1981) *Introduction to Montague Semantics*. Reidel, Dordrecht.

Dummett, M (1981) *Frege: Philosophy of Language*. Duckworth, London.

van Dijk, T A (1976) (ed.) *Pragmatics of Language and Literature*. North Holland, Amsterdam.

Fillmore, C J (1971) Verbs of judging: an exercise in semantic description. In Fillmore, C J and Langendoen, D T (eds.) *Studies in Linguistic Semantics*. Holt New York.

Fodor, J D (1979) In defence of the truth value gap. In Oh, C-K and Dinneen, D A (eds.) *Syntax and Semantics II: Presupposition*. Academic Press, New York.

van Fraassen, B C (1966) Singular terms, truth-value gaps and free logic. *Journal of Philosophy*, 63: 481-495.

van Fraassen, B C (1969) Presuppositions, supervaluations and free logic. In Lambert, K (ed.) *The Logical Way of Doing Things*. Yale University Press, New Haven.

Furberg, M (1971) *Saying and Meaning*. Basil Blackwell, Oxford.

Gazdar, G (1979) *Pragmatics: Implicature, Presupposition and Logical Form.* Academic Press, New York.

Geach, P and Black, M (1980) *Translations from the Philosophical Writings of Gottlob Frege*. Basil Blackwell, Oxford.

Geach, P (1959) Russell on meaning and denoting. *Analysis*, 4: 69-72.

Green, G M (1989) *Pragmatics and Natural Language Understanding*. Laurence Erlbaum Associates Inc., New Jersey.

Grice, P (1975) Logic and conversation. In Cole, P and Morgan, J (eds.) *Syntax and Semantics*, vol. 3. Academic Press, New York .

Harnish, R (1976) Logical form and implicature. In Bever, T, Katz, J and Langendoen T (eds.) *An Integrated Theory of Linguistic Ability*. Crowell, New York.

Hawkins, J A (1978) *Definiteness and Indefiniteness*. Croom Helm/Humanities Press, London/Atlantic Highlands, New Jersey.

Hawkins, J A (1984) A Note on Referent Identifiability and Copresence. *Journal of Pragmatics*, 8.

Heim, I R (1982) *The semantics of definite and indefinite noun phrases.* PhD thesis, University of Massachusetts.

Herzberger, H. (1970) Truth and modality in semantically closed languages. In Martin, R., (ed.) *The Paradox of the Liar*. Yale University Press, New Haven.

Hudson, R A (1984) *Word Grammar*. Blackwell, Oxford.

Kamp, H (1981) A theory of truth and semantic representation. In Groenendijk, J. A. G., T. M. V. Janssen and M. B. V. Stokhof (eds.). *Formal Methods in the Study of Language*, 1: 277-322.

Kamp, H and Reyle, U (1993)

Karttunen, L (1973) Presuppositions of compound sentences. *Linguistic Inquiry*, 4: 169-193.

Karttunen, L (1974) Presupposition and linguistic context. *Theoretical Linguistics*, 1: 181-194.

Karttunen, L and Peters, S (1979) Conventional Implicature. In Oh, C-K and Dinneen, D A (eds.) *Syntax and Semantics II: Presupposition*. Academic Press, New York.

Keenan, E L (1972) On semantically based grammar. *Linguistic Inquiry*, 3: 413-461.

Keenan, E L (1973) Presupposition in natural logic. *The Monist*, 57: 344-370

Keenan, E and Falz, L M (1985) Boolean Systems for Natural Language. *Linguistics and Philosophy*, 4: 159-219.

Kempson, R M (1975) *Presupposition and the Delimitation of Semantics*. Cambridge University Press, Cambridge.

Kempson, R M (1977) *Semantic Theory*. Cambridge University Press, Cambridge.

Kiparsky, P and Kiparsky, C (1971) Fact. In Steinberg, D D and Jakobovits, L (eds.) *Semantics*. Cambridge University Press, London.

Kripke, S A (1980) *Naming and Necessity*. Oxford University Press, Oxford.

Ladusaw, W, A (1979) *Polarity Sensitivity as Inherent Scope Relations*. PhD Thesis, University of Texas.

Lakoff, R (1972) The pragmatics of modality. In Peranteau, P et al. (eds.) *Papers from the Eighth Regional Meeting of the Chicago Linguistics Society*: 229-247. University of Chicago, Chicago.

Levinson, S C (1983) *Pragmatics*. Cambridge University Press, Cambridge

Lewis, D (1968) Counterpart theory and quantified modal logic. *Journal of Philosophy* 65: 113-126.

Lewis, D (1973) *Counterfactuals*. Harvard University Press, Cambridge, Massachusetts.

Lewis, D (1986) *The Plurality of Worlds*. Basil Blackwell, Oxford.

Lyons, J (1975) Deixis as the source of reference. In Keenan, E L (ed) *Formal Semantics of Natural Language*. CambridgeUniversity Press, Cambridge.

Lyons, J (1977) *Semantics*. Cambridge University Press, Cambridge.

Martin, J. (1975) Karttunen on possibility. *Linguistic Inquiry*, 6: 339-341.

May, R (1985) *Logical Form: Its Structure and Derivation*. MIT Press, Cambridge, Mass.

McCawley, J (1981) *Everything that linguists have always wanted to know about logic but were ashamed to ask*. Blackwell, Oxford.

McMahon, A M S (1994) *Understanding Language Change*. Cambridge Universirt Press, Cambridge.

Meinong, (1904) In *Untersuchungen zur Gegenstands-theorie und Psychologie*. Leipzig.

Morgan, J On the treatment of presupposition in transformational grammar. In Binnick, R, Davison, A, Green, G and Morgan J (eds) *Papers from the fifth regional meeting*. Chicago Linguistics Society.

Nute, D (1975) Counterfactuals and the similarity of worlds. *Journal of Philosophy*, 72: 773-778.

Partee, B H, ter Meulen, A and Wall, (1990) *Mathematical Methods in Linguistics*. Reidel, Dordrecht.

Prince, E F (1981) Toward a taxonomy of given - new information. In Cole, P (ed.) *Radical Pragmatics*. Academic Press, New York.

Quine, W V (1953) *From a Logical Point of View*. Harvard University Press, Cambridge Massachusetts.

Radford, A (1990) *Syntactic Theory and the Acquisition of English Syntax*. Blackwell, Oxford.

Rescher, N (1969) *Many Valued Logic*. McGraw Hill, New York.

Russell, B (1905) On Denoting. *Mind*, 14: 479-493.

Russell, B (1923) Vagueness. *Australian Journal of Philosophy*, 1: 84-92.

van der Sandt, R (1988) *Context and presupposition*. Croom Helm, Beckenham.

Seuren, P (1985) *Discourse Semantics*. Blackwell, Oxford.

Smaby, Richard M., (1979). Ambiguous Coreference with Quantifiers. In Guenthner, F. and Schmidt, S J. *Formal Semantics and Pragmatics for Natural Languages*. Reidel, Dordrecht.

Sperber, D and Wilson, D (1979) *Relevance Communication and Cognition*. Basil Blackwell, Oxford.

Stalnaker, R C (1974) Pragmatic presuppositions. In Munitz, M and Unger, P K (eds.) *Semantics and Philosophy*: 197-204. New York University Press, New York.

Stoel-Gammon, C and Cooper, J A (1984) Patterns of early lexical and phononlogical development. *Journal of Child Language*, 11: 247-271.

Strawson, P F (1950) On referring. *Mind*, 59: 230-244.

Strawson, P F (1952) *Introduction to Logical Theory*. Methuen, London.

Strawson, P F (1954) A reply to Mr Sellars. *Philosophical Review*, 63: 216-231.

Strawson, P F (1964) Identifying reference and truth-values. *Theoria*, 30: 96-118.

Strawson, P F (1974) *Positions for Quantifiers*. In Munitz M K & Unger, P K (eds.) *Sematics and Philosophy*. New York University Press, New York.

Traugott, E (1982) From propositional to textual and expressive meanings: some semantic-pragmatic aspects of grammaaticalization. In Lehmann W P and Malkiel Y (eds.) *Directions for Historical Linguistics*. University of Texas Press, Austin.

Traugott, E (1989) On the rise of epistemic meanings in English: an example of subjectification in semantic change. *Language*, 65: 31-55.

Vendler, Z (1967) *Linguistics in Philosophy*. Cornell University Press, Ithaca, New York.

Werth, P (1993) Accommodation and the myth of presupposition: A view from discourse. *Lingua* 89: 39-95

Wilson, D (1975) *Presuppositions and Non-Truth Conditional Semantics*. Academic Press, New York.

Wittgenstein, L (1921) *Tractatus Logico-Philosophicus*. English translation by Pears, D F and McGuinness, B F (1961), Routledge/Kegan Paul, London.

Zimmerman, H-J (1985) *Fuzzy Set Theory and its Applications*. Nijhoff Junk, Dordrecht